GARBAGE IN, GARBAGE OUT – MEASURING AND PREDICTING INSUFFICIENT EFFORT RESPONDING IN SURVEY RESEARCH

ARIANE GAUTHIER

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Doctor of Philosophy (Ph.D.) degree in Psychology

School of Psychology
Faculty of Social Sciences
University of Ottawa

© Ariane Gauthier, Ottawa, Canada, 2026

Abstract

The validity and reliability of conclusions drawn from survey data are contingent upon the quality of responses provided by participants. When analyzing survey data, researchers assume participants are engaged and exert a sufficient effort responding to the survey items. While it was initially believed that most survey research participants are attentive, a growing body of evidence suggests that up to 50% of participants (Francavilla et al., 2019; Meade & Craig, 2012) put forth an insufficient effort responding (IER). Although IER represents a significant threat to survey research integrity, important gaps persist in the literature pertaining to the detection and prediction of IER. On the one hand, a standardizable IER detection approach has yet to be developed. On the other hand, even if understanding the process underlying IER is crucial to prevent it, a robust and widespread examination of the predictors of IER remains to be conducted. This thesis sought to tackle the aforementioned gaps across six studies divided into three manuscripts.

In *Manuscript 1*, a novel standardizable IER detection approach (i.e., multidimensional insufficient effort responding detection approach; mIERda) was introduced and its criterion validity was examined across a series of four simulation studies. Building on *Manuscript 1*, *Manuscript 2* relied on a prospective design to further the validation of the mIERda and assess its construct validity using 18 datasets ($N = 5014$). Lastly, drawing on 23 datasets ($N = 2216$) collected over the course of two years, *Manuscript 3* examined the contextual (i.e., survey length, study topic, time of the semester) and personal (i.e., personality, global motivation, motivation for research participation) determinants of IER using a prospective longitudinal design. Taken together, findings arising from this thesis provide evidence of the criterion (*Manuscript 1*) and construct (*Manuscript 2*) validity of the mIERda. Results from *Manuscript 3*

demonstrate a limited contribution from contextual determinants and suggest that IER is mostly driven by personal determinants. Results also support that global autonomous motivation acts as a protective factor against IER at the start of the survey, whereas amotivation for research is tied to greater odds of IER throughout the survey. These findings provide meaningful insight to guide recommendations pertaining to IER detection, survey design, and motivational interventions with the overarching goal of lessening the prevalence of IER and its impact on survey research.

Keywords: insufficient effort responding, careless responding, self-determination theory, survey research, survey methodology, data quality

“J’accepte la grande aventure d’être moi”

Simone de Beauvoir

Acknowledgements

When I first started my Ph.D., it was important for me to work on something that would feel meaningful and from which I would draw a sense of pride. Years later, reflecting on this work, I feel proud not only of the research presented in my thesis, but also of the person I became along the way. I, however, would not be able to say as much without those who have supported me throughout my doctoral journey.

To Dr. Luc Pelletier, my supervisor. Thank you for seeing the potential in me back in 2019 and for taking me as your last doctoral student. Working with you has helped me grow both as a professional and as an individual. This experience has shaped me into the person I am today.

I extend my gratitude to the members of my thesis committee, Dr. Krystelle Shaughnessy, Dr. Serena Corsini-Munt, Dr. Jean-François Bureau, and Dr. Richard Koestner, for their interest, insight, and guidance throughout this project.

I also want to thank Dr. Chris Gravel without whom this thesis would not have been possible. Thank you for providing me with the methodological tools and knowledge I needed to turn my ideas into reality.

À François, Meredith, Bianca, Dominique, Geneviève, Isabelle et Stéphanie, merci d'avoir mis du soleil, chacun.e. à votre façon, dans mon parcours universitaire. Je chérirai toujours nos fous rires et les moments que nous avons passés ensemble.

À mes parents. Merci de m'avoir inculqué l'importance d'être rigoureuse et d'aller jusqu'au bout de mes ambitions. À ma fratrie : Élodie, Simon, Livia, Alexandre, Charles et Amélie qui, malgré la distance, m'inspirent tous les jours à être une meilleure personne. Un merci tout particulier à Élodie. Merci d'être toujours là pour moi lorsque j'ai besoin d'une oreille

attentive. Tu es l'épitome même d'une petite sœur exemplaire et je suis si chanceuse de t'avoir dans ma vie.

À Roseline et Marjolaine. Merci de m'avoir accueilli à bras ouverts dans votre quotidien et de me rappeler, chacune à votre façon, que la vie ne s'arrête pas au travail. Vous êtes magnifiques et je vous aime. Un merci particulier à Roseline, merci pour tous les magnifiques dessins que tu m'as offerts en guise d'encouragement.

À JF. Les mots me manquent pour t'exprimer à quel point ton amitié et ton mentorat ont transformé mon expérience doctorale. Ta bienveillance, ton ouverture ainsi que ta confiance en ma vision et mes compétences ont été un véritable pivot dans ma trajectoire professionnelle. Je ne pourrai jamais te remercier assez pour tout ce que tu as fait pour ma carrière. Pour l'instant, je me contenterai donc de te dire ceci: merci. Merci, tellement.

Finalement, un merci tout spécial à mon conjoint, Simon. Merci d'avoir cru en moi lorsque je n'arrivais plus à le faire. Merci pour toutes les petites attentions, ton support indéfectible, ton expertise bienveillante et ton accueil inconditionnel. Merci d'avoir été, et de continuer d'être, mon roc contre vents et marées. Merci d'avoir tenu le phare pour l'ensemble de la famille lors des (interminables!) mois ayant précédé le dépôt de ma thèse. Tous les jours, je me dis que je suis drôlement chanceuse de partager mon quotidien avec une personne aussi attentionnée et dévouée que toi. Je n'aurais pas pu espérer avoir une meilleure personne à mes côtés pendant cette période de ma vie. Merci pour tout. Je t'aime et j'ai hâte de débuter ce nouveau chapitre avec toi.

In memoriam de Poséidon, mon fidèle compagnon félin qui m'a accompagnée du tout début jusqu'à la toute fin de mes études universitaires. Je me sens profondément privilégiée de savoir que tu as consacré les derniers mois de ta vie à me tenir compagnie, lové entre ma thèse et moi.

Ton ronronnement est la trame sonore de cette thèse.

Preface

Ethics Approval

The research presented in the present thesis received the approval of the University of Ottawa Research Ethics Board (#H-04-25-11034).

Statement of Funding

The work presented in the present thesis was supported by master's and doctoral scholarships from the Social Sciences and Humanities Research Council of Canada, les Fonds de recherche du Québec, and the Government of Ontario. The research presented in the present thesis was supported by a grant (SSHRC IG grant #435-2021-0344) from the Social Sciences and Humanities Research Council « *Detecting, predicting, and reducing insufficient effort responding: A motivational approach* » to the thesis supervisor, Dr. Luc Pelletier.

Statement of Co-Authorship

This thesis is composed of five chapters: a general introduction, three manuscripts, and a general discussion. The general introduction (Chapter I) provides an overview of the theoretical background that guided each portion of the present research program, the three manuscripts present the findings of six studies (Chapters II – IV), and the general discussion (Chapter V) presents an integration of the findings. I, Ariane Gauthier, am responsible for the conceptualization, methodology, data curation, investigation, data analysis, and writing of this thesis, and I am the primary author on all manuscripts. My thesis supervisor, Dr. Luc Pelletier, provided feedback at all stages of the thesis, and is the second author on all manuscripts.
Dr. Chris Gravel also provided computational resources and feedback on *Manuscript 1*, and is the third author of this manuscript.

Table of Contents

Abstract ii
Acknowledgements v
Preface viii
Ethics Approval viii
Statement of Funding viii
Statement of Co-Authorship viii
List of Tables xiii
List of Figures xvii

CHAPTER I : GENERAL INTRODUCTION 1

Insufficient Effort Responding 4

Impact of Insufficient Effort Responding on Research Results 4

Detecting Insufficient Effort Responding 5

Attention Checks and Self-Reported Measures 6
Paradata 9
Data-Driven Approach 10
Profiles of Survey Participants 16
Why Use Various Indicators of IER? 17
Limitations of IER Detection Approaches 18

Predicting Insufficient Effort Responding 20

Contextual Determinants of IER 20
Personal Determinants of IER 23
Limitations of the Research on the Determinants of IER 26

Self-Determination Theory 27

Hierarchical Model of Intrinsic and Extrinsic Motivation 30

The Present Dissertation 33

CHAPTER II : SORTING THE GOOD FROM THE MIERDA – INTRODUCING AND TESTING THE PERFORMANCE OF THE MULTIDIMENSIONAL INSUFFICIENT EFFORT RESPONDING DETECTION APPROACH 35

Abstract 36

Introduction 38

Impact of IER 38
Identifying IER Respondents 39
Multidimensional Insufficient Effort Responding Detection Approach (mIERda) 47
Present Research 50

Study 1a 51

Methods 51
Plan of Analysis 55
Performance Measures 56

Results and Discussion 57
Summary 61

Study 1b 63

Method 63
Results and Discussion 64
Summary 64

Study 2 65

Methods 65
Results and Discussion 66
Summary 67

Study 3 67

Methods 68
Results and Discussion 69
Summary 70

General Discussion 71

Main Findings 72
Rethinking our Approach to IER Detection 77
Limitations & Future Research 78
Conclusion 80

References 81

CHAPTER III : MULTI-STUDY EXAMINATION OF THE NOMOLOGICAL VALIDITY OF THE MULTIDIMENSIONAL INSUFFICIENT EFFORT DETECTION APPROACH (MIERDA 103

Abstract 104

Introduction 105

A Novel Approach to IER Detection: the Multidimensional Insufficient Effort Responding Detection Approach (mIERda) 106
Nomological Network of Insufficient Effort Responding 107
Limitations of Existing Research 110
Current Study 111
Hypotheses 112

Method 113

Participants 113
Procedures and Measures 113

Results 117

Preliminary Analyses 117
Descriptive Statistics 118
Convergent Validity 118
Discriminant Validity 119

Predictive Validity 119
Concurrent Validity 120
Known-Groups Validity 121

Discussion 121

Validity of the mIERda 122
The mIERda Throughout the Survey 127
The mIERda vs. Other Indicators of IER 128
Limitations and Future Research 130
Conclusion 131

CHAPTER IV : JOINT EXPLORATION OF THE CONTEXTUAL AND PERSONAL DETERMINANTS OF INSUFFICIENT EFFORT RESPONDING ACROSS TIME – A SELF-DETERMINATION THEORY PERSPECTIVE 155

Abstract 156

Introduction 158

Determinants of Insufficient Effort Responding 160
Revisiting the Role of Motivation in IER: The Self-Determination Theory Perspective 165
Limitations of Past Research 171
Present Research 175
Hypotheses 175

Method 177

Participants 177
Procedure 177
The Multidimensional IER Detection Approach 179
Substantive Measures 180

Results 187

Preliminary Analyses 187
Examination of the Determinants of IER 188

Discussion 196

Effect of Motivation 196
Personality 198
Time of the Semester 200
Survey Length 201
Survey Topic 201
Covariates 202
Limitations and Future Research 203
Conclusion 204

References 205

CHAPTER V : GENERAL DISCUSSION 227

Summary of Thesis Objectives 227
Summary of Main Findings 228

Implications of the Present Thesis 236
Limitations of the Thesis and Directions for Future Research 245
Conclusion 251
References 253
APPENDIX A 277
APPENDIX B 282
APPENDIX C 288
APPENDIX D 311
APPENDIX E 324
APPENDIX F 329

List of Tables

CHAPTER II – Sorting the Good from the mIERda – Introducing and Testing the Accuracy of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

Table 2.1: *Overview of Simulated Scale Characteristics* ______ *92*

Table 2.2: *Overview of the Data-Driven IER Indicators* ______ *93*

Table 2.3: *Summary of the Performance Measures* ______ *94*

CHAPTER III – Multi-Study Examination of the Construct Validity of the Multidimensional Insufficient Effort Detection Approach (mIERda)

Table 3.1: *Aggregated Socio-Demographic Information* ______ *145*

Table 3.2: *Summary of Constructs and Measures for the Validation of the mIERda* ______ *146*

Table 3.3: *Aggregated Descriptive Statistics of the Numerical Validation Variables* ______ *149*

Table 3.4: *Descriptive Statistics of the Categorical Validation Variables* ______ *150*

Table 3.5: *Pooled Correlation Estimates between the Multidimensional Insufficient Effort Responding Detection Approach Classification at the Start and End of the Survey and Validation Variables.* ______ *151*

Table 3.6: *Aggregated mIERda Overlap/Sensitivity with Attention Checks and Self-Reported Items* ______ *152*

Table 3.7: *Prevalence of IER as per the Multidimensional Insufficient Effort Responding Approach by Datasets and by Research Programs* ______ *153*

CHAPTER IV – Joint Exploration of the Contextual and Personal Determinants of Insufficient Effort Responding Over Time – A Self-Determination Theory Perspective

Table 4.1: *Aggregated Socio-Demographic Information* ______ *216*

Table 4.2: *Correlation Matrix between Variables of Interest* ______ *217*

Table 4.3: *Stepwise Multilevel Logistic Regression Results* ______ *218*

Table 4.4: *Stepwise Model Comparison of Effects* ______ *221*

Table 4.5: *Contrasts Across Hierarchical Levels of Motivation and Across Categorical Variables for Model 3* ______ *222*

APPENDIX C

Table C-3.1: *Descriptive Information of Research Programs* ______ *289*

Table C-3.2: *Subject Pool Opening and Closing Date by Semester* ______ *290*

Table C-3.3: *Summary of Attention Checks by Research Programs and Datasets* ______ *291*

Table C-3.4: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 1* ______ *292*

Table C-3.5: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 2* ______ *293*

Table C-3.6: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 3* ______ *294*

Table C-3.7: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 5* ______ *295*

Table C-3.8: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 7* ______ *296*

Table C-3.9: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 8* ________________ *297*

Table C-3.10: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 9* ________________ *298*

Table C-3.11: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 10* _______________ *299*

Table C-3.12: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 11* _______________ *300*

Table C-3.13: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 12* _______________ *301*

Table C-3.14: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 13* _______________ *302*

Table C-3.15: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 17* _______________ *303*

Table C-3.16: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 18* _______________ *304*

Table C-3.17: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 19* _______________ *305*

Table C-3.18: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 20* _______________ *306*

Table C-3.19: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 21* _______________ *307*

Table C-3.20: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 22* 308

Table C-3.21: *Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 23* 309

Table C-3.22: *mIERda Overlap/Sensitivity with Attention Checks and Self-Reported Items by Research Programs and Datasets* 310

APPENDIX D

Table D-1: *Summary of Measures by Research Programs and Datasets – Manuscript 2 & 3* 312

APPENDIX E

Table E-4.1: *Descriptive Information of Research Programs – Manuscript 3* 325

Table E-4.2: *Subject Pool Opening and Closing Date by Semester* 326

Table E-4.3: *Socio-Demographic Information by Research Programs – Manuscript 3* 327

Table E-4.4: *Frequency and prevalence of IER as per the Multidimensional Insufficient Effort Responding Approach by Study Topic – Manuscript 3* 328

List of Figures

CHAPTER II – Sorting the Good from the mIERda – Introducing and Testing the Accuracy of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

Figure 2.1: *Study 1a – Area under the Curve of the IERI with and without a Cutoff* ______ *95*

Figure 2.2: *Study 1a – Area under the Curve of Additional IERI* ______ *96*

Figure 2.3: *Study 1a – Balanced Accuracy of the Sequential Approach and the mIERda Across Scales and Prevalence Rates* ______ *97*

Figure 2.4: *Study 1a – Balanced Accuracy of the Sequential Approach and the mIERda Across Scales and Severity Levels* ______ *98*

Figure 2.5: *Study 1b – Balanced Accuracy of the mIERda Across Scales, Prevalence and Severity Rates* ______ *99*

Figure 2.6: *Study 2 – Balanced Accuracy of the mIERda for Mixed Respondents Across Prevalence and Severity Rates for the HEXACO (Scale 3)* ______ *100*

Figure 2.7: *Study 3 – Balanced Accuracy of the mIERda for Mixed Respondents Across Prevalence and Severity Rates for the ECR (Scale 1) and BFI (Scale 2)* ______ *101*

Figure 2.8: *Study 3 – Sensitivity and Specificity of the mIERda for Mixed Respondents Across Prevalence, Severity Rates and Levels of Screening* ______ *102*

CHAPTER III – Multi-Study Examination of the Construct Validity of the Multidimensional Insufficient Effort Detection Approach (mIERda)

Figure 3.1: *Hypothesized Nomological Network of the multidimensional Insufficient Effort Responding detection approach (mIERda)* ______ *154*

CHAPTER IV – Joint Exploration of the Contextual and Personal Determinants of Insufficient Effort Responding Over Time – A Self-Determination Theory Perspective

Figure 4.1: *Data Collection Process* ______ *224*

Figure 4.2: *Start of the Survey - Effect of the Numerical Contextual and Personal Determinants on Odds of IER* ______ *225*

Figure 4.3: *End of the Survey - Effect of the Numerical Contextual and Personal Determinants on Odds of IER* ______ *226*

CHAPTER I : General Introduction

Survey research has roots in centuries of history. The first survey can be traced back to Charles Booth in the late 1800s. Booth took it upon himself to examine the living conditions of the poorest in London (Ornstein, 2013). Although drastically different from modern surveys, his work not only led to significant social reforms but also laid the foundation for empirical research in the social sciences. Even if the apparition of census, psychological and intelligence testing followed his work in the 1920s, it was not until the late 1930s that surveys were first used in academic and psychological research (Lazarsfeld & Fiske, 1938; Ornstein, 2013). Ever since, social scientists have never stopped using surveys to collect descriptive data, make inferences about populations, and propose new theoretical models (Converse, 2009).

When researchers collect data using surveys, the integrity of said data depends on participants putting forth a sufficient effort in reading, understanding, and responding to the survey. According to Ornstein (2013), one of the main reasons Booth's contribution was so meaningful is that, through his work, he accurately portrayed the experience of the population he surveyed. Similarly, as researchers conduct analyses and draw conclusions from data collected through surveys, their end goal is (or at least should be) to accurately depict their population of interest. When collecting data through surveys, researchers consequently assume that their participants share a similar willingness to be truthful, accurate, and provide responses resulting from an effortful responding process. This assumption has, however, been challenged by results arising from a growing body of research.

According to numerous researchers, a nonnegligible number of participants provide unreliable data and put forth an insufficient effort responding (IER) when taking part in survey research (Bowling et al., 2016; Huang et al., 2012; Huang, Liu, Bowling, et al., 2015). Unless

they adequately address IER in their samples, researchers end up conducting their analysis on partially spurious, thus potentially meaningless data that, if left unaddressed, can affect the reliability of their research results (e.g., Arias et al., 2020, 2022; Huang, Liu, & Bowling, 2015; Kam & Meyer, 2015; Oppenheimer et al., 2009).

Given the potentially pervasive impact of IER, researchers have taken interest in improving its detection. However, on a larger scale, identifying IER is not as easy as it seems. IER is not well known, nor is it well defined by the scientific community. While some researchers have proposed recommendations (Curran, 2016; Ward & Meade, 2023), we still do not have a clear idea of how it should be measured, when each indicator should be used to identify participants engaging in IER, and which cut-offs should be used for each of these indicators. Addressing these problems, or finding an approach to circumvent them, is critical if we want to identify IER and report it. In the absence of clear guidelines or a standardized approach, researchers have been picking the indicators and cutoffs they saw fit, making it impossible to draw a clear picture of the phenomenon. As long as the operationalization of IER differs, it remains complex not only to estimate the prevalence of IER but also to shed light on the circumstances under which it occurs.

As a result, it remains unclear if IER is related to specific study topics, or other contextual characteristics (e.g., survey length, topics, time of the semester, being rewarded or not for participating, etc.). Although past research has investigated the impact of some of these contextual characteristics (e.g., Bowling et al., 2021, 2022; Brower, 2018; Cheung, 2019; Galesic & Bosnjak, 2009; Gibson & Bowling, 2020), important limitations linger. Firstly, the effects of these contextual characteristics on IER are often investigated individually; thus, results drawn from these studies do not simultaneously control for other study's characteristics (i.e., survey

length, study topics, time of the semester) which makes it impossible to establish causality. For instance, it is unknown whether the effect of survey length on IER remains significant if we account for study topics. Additionally, researchers each use different indicators to detect IER, which means that IER is operationalized differently in each study. It thus remains unknown whether the differences in determinants and prevalence observed across studies are substantial or result from methodological artefacts. To increase the quality of survey data, it is important for these methodological limitations to be addressed as it will allow us to identify the contextual determinants which may foster IER, and design interventions to decrease their impact.

Besides investigating the contextual determinants of IER, understanding which, when, and how participants' characteristics can impact their engagement in IER is also crucial to getting a full picture of the phenomenon. For example, although contextual determinants seem to have an impact on IER, it is plausible that these determinants do not predict IER as strongly when we account for the effect of participants' motivation, personality and other personal characteristics (i.e., gender, age, ethnicity). The idea that personal determinants might affect engagement in IER aligns with its definition as a lack of motivation to engage in research activities (Huang et al., 2012). However, even if motivation (or lack thereof) is at the very core of the definition of IER, research has yet to robustly establish its relationship with IER. Although results from past research offers initial evidence of motivation's role in IER, none have considered motivation alongside other personal determinants (e.g., personality), contextual determinants (e.g., survey length), different study topics, and a prospective or longitudinal approach. As a result, it is unknown to us whether a lack of motivation truly underlies participants' engagement in IER, or if other variables are better predictors of IER. Overall, the literature not only highlights a need to standardize and facilitate the detection of IER, but also to

thoroughly examine the determinants of IER using a robust methodological and analytical approach.

In the face of these limitations, the goals of this thesis were to (a) improve the detection of IER and to (b) examine the contextual (i.e., topic, length, time of semester) and personal (e.g., motivational, personality) determinants of IER. In doing so, the goal was to better understand how to assess IER as well as what promotes its occurrence, which will provide us with meaningful insight to eventually prevent it.

Insufficient Effort Responding

In survey research, IER is defined as a problematic response style characterized by low or little motivation to follow survey instructions, interpret item content, and provide accurate responses (Huang et al., 2012; Meade & Craig, 2012). As they engage in IER, participants are disengaged, inattentive, and do not provide accurate responses. Instead of providing item content-responsive responses, IER participants will provide responses that are non-responsive to the item content. As exposed below, when left unaddressed, these participants represent a considerable threat to data quality (DeSimone & Harms, 2018), thus to the reliability and validity of results and conclusions drawn from survey research.

Impact of Insufficient Effort Responding on Research Results

It is currently estimated that IER participants represent between 3% and 50% of research samples, in rates varying greatly between studies (Francavilla et al., 2019; Maniaci & Rogge, 2014; Meade & Craig, 2012). Whilst IER represents a considerable proportion of samples, it is also known that IER often goes undetected, therefore unaddressed, by typical data cleaning procedures (Curran, 2016). These findings are concerning, as it has been demonstrated that a proportion of 10% IER participants suffices to increase the rate of type I and II errors (Huang,

Liu, & Bowling, 2015). It is suspected that IER leads to the publication of spurious results, which subsequently limits our ability to reproduce these "findings" (Curran, 2016).

For example, results from additional studies suggest IER can distort the correlation between variables of interest (Huang, Liu, & Bowling, 2015; Kam & Meyer, 2015), threaten factorial analysis results (Arias et al., 2020; Huang, Liu, & Bowling, 2015; Kam, 2019; Woods, 2006), and diminish statistical power, construct validity and effect sizes (Kam & Meyer, 2015; Maniaci & Rogge, 2014; Oppenheimer et al., 2009). While psychology finds itself in the midst of a replication crisis, with only 39% of findings and 47% of effect sizes being reproduced (Aarts et al., 2015); the survey-reliant field of social psychology finds itself particularly impacted with replication rates varying between 20% and 45% (Aarts et al., 2015; Schimmack, 2020). Although the exact cause of this finding remains unknown, IER is suspected of contributing to this phenomenon (Curran, 2016). Overall, this supports the relevance of addressing this phenomenon. Doing so, however, requires a valid and accurate IER detection approach.

Detecting Insufficient Effort Responding

Given the known pervasive impact of leaving IER unaddressed, some researchers aware of that threat have taken an interest in detecting IER in their samples. As a result, multiple indicators of IER have been introduced in the literature. Researchers have thus been relying on various indicators of IER, such as questions meant to test participants' attention (Meade & Craig, 2012; Oppenheimer et al., 2009), analysis of any suspicious response pattern (Johnson, 2005), or examination of completion time (Bowling et al., 2023) to identify IER participants. These different indicators have been extensively summarized in the literature (Curran, 2016; Desimone et al., 2015; Huang et al., 2012; Huang, Liu, & Bowling, 2015; Ward & Meade, 2023).

These IER indicators, which are being presented in greater detail below, can be clustered into three broad categories: (1) attention checks and self-reported measures (e.g., instructed and bogus items), (2) paradata (e.g., completion time) and (3) data-driven indicators (e.g., long string index, Mahalanobis distances, individual variability, person-total correlation).

Attention Checks and Self-Reported Measures

Attention checks and self-reported items seek to identify instances of IER by inviting participants to either (a) provide a specific answer or (b) report on the quality of the answer they provided in the survey. This family of indicators encompasses instructed, bogus, and self-reported items.

Instructed Items

Instructed items (e.g., "*Please respond Slightly Agree to this item*", "*Please answer 7 to this item"*) are items added to a survey to which the expected response is specified within the item content. Participants respond to these items using a Likert scale. These items are added inconspicuously amongst the other survey items as the goal is not to draw participants' attention, but to check if they were paying attention to the survey at that specific moment. To get a glimpse of participants' attention throughout their study, researchers can add instructed items across the survey. Participants failing to provide the expected response are considered potentially IER. Meade & Craig (2012) suggest that one instructed or bogus item should be added every 50 to 100 items. Curran (2016) suggests adding at least one instructed item at the start, mid, and end of the survey. The main advantage of instructed items relies on the straightforwardness of the indicator: unlike bogus items, instructed items leave no room for ambiguity.

Bogus Items

Similarly, bogus items are added to a survey, among its items, to identify inattentive participants. Bogus items are statements with which it is impossible or very unlikely to agree with such as, "*I was born on February 30th*" or "*I have never used a computer*" (Beach, 1989). Participants respond to these items using a Likert scale. Although very similar to instructed items, bogus items differ as the expected response to the item is implied rather than being specified in the item content. These items are responded to using a Likert-scale.

From an applicability standpoint, bogus items are trickier to score than instructed items as it is possible for bogus items to have multiple correct answers. For example, responding "Very false" or "False" to "I eat cement occasionally" on a 5-point scale ranging from "1–Very false" to "5–Very true" are two equally valid answers and participants picking either of these response choices could be considered "attentive". Since more than one correct answer may exist, bogus items are less sensitive to IER than instructed items. Statistically, a participant responding completely randomly to a survey containing a bogus item (for which there are two known correct answers) has a 2 out of 5 (40%) probability on a 5-point scale, and a 2 out of 7 (28%) probability on a 7-point scale of randomly picking a correct answer and not being caught by the item. This probability is cut in half for an instructed item. A past study has also shed light on another limitation of bogus items by postulating that these items could confuse some attentive participants, leading to their incorrect classification as IER (Curran & Hauser, 2019). It is posited that since survey participants do not expect trick questions when participating in research, they might agree with the statement if a part of it is true, even if it does not entirely capture their experience (Curran & Hauser, 2019), leading to their false classification as IER. While these items reportedly work as well as other IER indicators, it is generally recommended that

researchers use instructed items over bogus items in an IER identification context (Curran, 2016), and additionally rely on other indicators, such as self-reported items.

Self-Reported Items

Found at the very end of the survey, the three self-reported data quality items developed by Meade & Craig (2012) are meant to capture participants' self-reported levels of commitment to the survey. Precisely, the proposed items capture the self-perceived level of (1) effort and (2) attention paid to the survey, and (3) the overall quality of the responses they provided throughout the study.

Using a 5-point Likert scale, participants rate how much effort they put in the study (i.e., self-reported effort; "*I put forth____ effort towards this study")* and how much attention they paid to the study (i.e., self-reported attention; "*I gave this study ____ attention*"). Then, participants indicate using a "Yes/No" statement whether they believe researchers should include the data they provided in their analysis (i.e., self-reported data inclusion; "*In your honest opinion, should we use your data?*"; refer to Meade & Craig (2012) for the exact protocol). Since these items capture a self-assessment of participants' level of commitment to the study, these items should not be positioned amongst other survey items and should rather be positioned at the end of the survey, to grab distracted participants' attention. Since this method depends on participants' truthfulness, it is thus vulnerable to IER, or other biases such as social desirability (DeSimone et al., 2015). While useful in identifying some IER participants, results from past research support that instructed, bogus and self-reported data quality/inclusion items have a very limited sensitivity or specificity to IER (Niessen et al., 2016). As such, researchers are also invited to consider other approaches such as paradata and data-driven indicators.

Paradata

In contrast with attention checks and self-reported items, which rely on direct observations to detect IER, the second category of IER indicators relies on paradata to capture IER. This approach relies on the analysis of participants' page- or survey-level completion time (Bowling et al., 2023; Meade & Craig, 2012; Ulitzsch et al., 2022), and more recently, on log data (Kroehne & Goldhammer, 2018), or keystrokes and mouse clicks (Kieslich & Henninger, 2017; Olson & Parkhurst, 2013). Overall, the assumption of using paradata is that if IER participants are or become disengaged, they should interact differently with the survey than attentive participants. This disengagement should thus translate into shorter or longer completion times, differences in keystrokes, or number of mouse clicks. Besides being useful for IER detection, these indicators have also been useful to detect bots (i.e., a program or script used to fill surveys with fake responses; Buchanan & Scofield, 2018), especially since these indicators can be programmed to respond correctly to attention checks.

While other paradata indicators (i.e., log data, keystrokes, mouse clicks) are relatively more recent, reliance on completion time does not date from yesterday. Completion time is a readily available metric for most online surveys, which makes it an easy-to-implement IER indicator. Researchers looking at completion time to identify IER have used various cut-offs such as a 2.5 interquartile range below the first quartile or above the third quartile (Funke, 2016); one- or two-standard deviations above or below mean completion time (Ashley & Shaughnessy, 2021; Heerwegh, 2003); a global completion time shorter than two seconds by item (Huang et al., 2012); a completion time lower than the first percentile (Gummer & Roßmann, 2014) or lower than the fifth percentile (Harms et al., 2017). Participants with a completion time outside the interval pre-established by the researcher would be deemed as IER. However, as the mean

and standard deviation are all impacted by the sampling distribution (Leys et al., 2013; Miller, 1991), cut-offs dependent on the normality of the distribution might fail to capture IER participants in the presence of extreme outliers (e.g., a participant keeping the survey open for days before submitting their response leading to an extremely long completion time consequently skewing the distribution). This approach also tends to have low statistical power in small samples (Cousineau & Chartier, 2010).

In addition to these limitations, it is worth bearing in mind that paradata indicators can only be derived from surveys administered from a computer. Furthermore, attention checks and paradata alike cannot be obtained retroactively, which might prevent their implementation in certain instances (e.g., archival data, paper-and-pencil data collection, omission by the researcher). In such instances, the data-driven approach might constitute a more appropriate and generalizable approach.

Data-Driven Approach

The data-driven approach refers to the application of statistical analyses to facilitate the identification of IER, instead of adding supplementary items or features (Karabatsos, 2003). The main advantage of this approach is that no preparation is needed before data collection since these indicators can solely be computed and analyzed afterwards. This means that this approach can be used a posteriori by researchers who want to capture IER in their sample. The data-driven approach comprises classic indicators of IER such as Mahalanobis distance (Meade & Craig, 2012), individual variability (Marjanovic et al., 2015), psychometric antonyms and synonyms (Meade & Craig, 2012), semantic synonyms and antonyms (Goldberg & Kilkowski, 1985), odd-even consistency (Huang et al., 2012; Meade & Craig, 2012), person-total correlation (Karabatsos, 2003), long string index (Johnson, 2005), and other more recent indicators (i.e.,

response coherence and response reliability; Dupuis et al., 2015, 2018). The information provided by each of these indicators is discussed below.

Mahalanobis Distance

Mahalanobis distance (MD; Mahalanobis, 1936) is an analysis used to highlight multivariate outliers. MD represents the multivariate distance of a case from the other cases observed in the sample (Tabachnick & Fidell, 2007). MD yields the probability of each case of being a multivariate outlier. Generally, a cut-off of p <.001 is used to classify a case as an outlier. However, some suggest a cutoff of $p < .05$ might offer greater sensitivity while remaining specific (Goldammer et al., 2020). While such cutoffs are commonly used, it is worth noting that the optimal cutoff for Mahalanobis distance, like any other IER indicator, is bound to change depending on the survey and sample characteristics (Falk et al., 2025; Ilagan & Falk, 2023). It is suggested in the literature that if participants are disengaged or do not pay attention to the survey items, their answers should widely differ from the population and they should thus get flagged as multivariate outliers (Curran, 2016). Although useful in identifying response patterns with high levels of variability (i.e., random responding), results from past research state that its sensitivity to IER response patterns associated with less variability (i.e., centered patterned responding) remains limited (Meade & Craig, 2012).

Intra-Individual Variability

The purpose of looking at individual variability is to flag inconsistent participants. Intra-individual reliability (Marjanovic et al., 2015) looks at the degree to which each respondent strays from their personal midpoint across items. High intra-individual variability can suggest the presence of random responses, which can be suggestive of a greater risk of IER. Curran (2016) argues that this indicator should be used cautiously as it requires a deeper understanding of

scores distribution and is insufficient, in its current shape, to detect IER. As individual reliability is often computed at a survey level instead of a scale level, it is sensitive to any changes in the survey design, such as Likert scale length. It is also worth noting that this indicator is also independent from the responses provided by the other participants in the sample, which makes it useful in conditions where a high prevalence of IER is expected.

Psychometric Synonyms and Antonyms

The purpose of psychometric antonyms and synonyms is to flag participants who provide inconsistent answers by creating psychometric pairs based on the sample's item correlations (Johnson, 2005; Meade & Craig, 2012). Within a provided sample, the psychometric synonym indicator searches for highly and positively correlated pairs of items, whereas the psychometric antonym indicator seeks to identify highly and negatively correlated pairs of items. In both instances, a value between -1 and 1 is returned for each participant.

Although informative, there are no clear cut-off scores to (a) generate pairs of items and (b) categorize participants as IER. Barnes (2018) has suggested using a correlation of -.60 (antonyms) or .60 (synonyms) to identify the pairs of highly correlated items. It is proposed that participants should be deemed IER if the absolute correlation is weakly positive or below 0 for psychometric synonyms (Barnes, 2018) and weakly negative or above zero for psychometric antonyms. However, the current literature does not offer sufficient support to justify this cut-off, making scoring subjective. Furthermore, this approach has reportedly been performing poorly, especially compared to other IER indicators (Barnes, 2018). Given the known effect of IER on inter-item correlations (Huang, Liu, & Bowling, 2015; Kam & Meyer, 2015), it is also likely that this indicator might lack reliability when the prevalence of IER is high in a sample.

Semantic Synonyms and Antonyms

This indicator is computed by creating pairs of semantically similar (synonyms; e.g., “I feel attentive” vs. “I feel alert”) and dissimilar (antonyms; e.g., “I feel proud” vs. “I feel ashamed”) items before any type of data analysis (Goldberg & Kilkowski, 1985). The researcher creates pairs they believe should be strongly correlated based on previous conceptual or empirical knowledge. Meade & Craig (2012) recommend using thirty pairs of items. Once again, researchers have yet to establish standards as to what represents a strong (or weak) enough correlation between these items, as well as a standardizable cutoff to identify IER. Semantic antonyms and synonyms return, for each participant, a value that ranges between -1 and 1.

Odd-Even Consistency

The odd-even consistency indicator is obtained by computing the correlation between the odd and even items within a unidimensional scale (Huang et al., 2012; Huang, Liu, & Bowling, 2015; Jackson, 1977; Johnson, 2005; Meade & Craig, 2012). The correlation between the odd and even items on the scales contained within a survey is then computed. Doing so returns a value ranging from -1 to 1. While odd-even consistency is typically recommended for IER screening (Curran, 2016; Ward & Meade, 2023), results from previous research suggest that other indicators such as person-total correlation have better discriminative properties to distinguish between IER and attentive participants (Dupuis et al., 2018).

Person-Total Correlation

The person-total correlation (PTC) represents a participant’s response consistency compared to the rest of the sample. The PTC therefore depends on the responses provided by the other participants. To compute the PTC, the correlation between individual responses and sample-level mean responses is computed for each item. The PTC returns a value ranging

between -1 and 1 for each participant. A PTC closer to 1 suggests a strong and positive correlation between the participant's responses and the sample's responses, whereas a PTC closer to -1 suggests a strong and negative correlation between the participants' responses and the sample. It is stated in the literature that a low positive, or a negative PTC should be considered a sign of IER (Curran, 2016).

Long String Index

The long string (Johnson, 2005; McCrae & Costa, 2006) is an indicator used to flag overly consistent participants. The long string index can be computed either at the survey or scale level. The long string index measures the frequency at which a same answer has been given in a row. A common recommendation is that a participant should be classified as IER if they consistently provide the same answer for at least half (50%) of the scale's items (Huang et al., 2012). A common recommendation is that a participant should be classified as demonstrating IER if they consistently provide the same answer for at least half (50%) of the scale's items (Huang et al., 2012). However, we suggest that this cut-off might not be optimal for shorter unidimensional measures (e.g., a 6-item scale) or for longer multidimensional measures (e.g., a 150-item scale). It remains unclear whether the long string index should be analyzed at a scale or survey level. It also remains unclear whether scoring as long string on one scale is enough to be deemed IER or whether this pattern should also be observed on multiple scales.

Response Coherence

The response coherence (Dupuis et al., 2015, 2018) indicates the extent to which the responses provided by a participant are predictable. It is computed by running a series of principal component analyses and varimax rotations until strict orthogonality is obtained. Doing so allows it to untangle the variance tied to the survey items from the variance tied to the

participant and to obtain a measure of response coherence. The response coherence has a possible range of [0, 1], where a score nearing 0 is indicative of a low level of coherence between the participant's responses and the expected factorial structure, which can suggest IER. While this indicator has yet to be widely adopted, results from previous research support that it had greater discriminative potential (i.e., to distinguish between IER and attentive participants) than other data-driven indicators of IER (Dupuis et al., 2018).

Response Reliability

The response reliability (Dupuis et al., 2015, 2018) provides a measure of the reliability of the responses provided by a participant. To obtain the response reliability, one must compute a series of principal component analyses and varimax rotations until strict orthogonality is achieved. Then, the resulting response strategy vector is divided into two parallel vectors. This is done by identifying pairs of closest items using the minimum Euclidean distance. The response reliability can then be obtained by computing the correlation between these parallel vectors and applying the Spearman-Brown correction. This indicator returns a value between -1 and 1 for each participant. A value closer to 1 is indicative of greater response reliability, but a cut-off value has yet to be established (Dupuis et al., 2015, 2018).

As displayed above, there are numerous types of indicators reported in the literature, and each of them highlights IER through distinct mechanisms. Such diversity is important because while all IER participants engage in IER, not every IER respondent engages in IER in the same way. Some participants engage in IER from the beginning to the end of the survey, while others might start off as attentive and become IER throughout the survey. Not only can IER participants differ in terms of onset, but they can also differ in terms of response patterns.

Profiles of Survey Participants

Using various IER indicators, research has previously highlighted three types of survey participants (Craig & Meade, 2012). Not only can we distinguish between attentive and IER participants, but the authors could also differentiate two profiles of IER participants: random and patterned IER participants (Craig & Meade, 2012). These attentive, random IER, and patterned IER participants differ from one another in the way they interact with the different IER indicators.

Attentive Participants

Compared to the two profiles of IER participants, attentive participants spend the most time completing the survey, miss fewer instructed and bogus items, have shorter long strings of response, higher consistency among psychometric synonyms, antonyms, and even-odd consistency indicators (Meade & Craig, 2012). Overall, they showed greater engagement with the survey.

Random IER Participants

This subgroup of IER participants provided responses to the survey items in a uniformly random manner. Concretely, the responses provided by these participants do not follow any clear pattern, and each response choice has approximately the same probability of being selected. Those participants spent less time on the survey, missed more instructed and bogus items than attentive participants, and had lower psychometric synonyms and odd-even consistency than attentive participants. Random participants also had shorter long strings of responses when compared to patterned participants (Meade & Craig, 2012).

Patterned IER Participants

These IER participants can be distinguished from random participants since their response strategy follows an observable pattern or trend. These participants can either provide invariant responses (e.g., providing an excessively long string of consecutive responses) or non-invariant responses (e.g., centered at the midpoint of the scale, focused on both ends of the Likert response scale, patterned). Patterned IER participants exhibited reduced survey completion times, overlooked numerous instructed and bogus items, and demonstrated moderately high psychometric and odd-even consistency compared to other participants. Compared to attentive and random IER participants, these participants were characterized by longer strings of repetitive responses (Meade & Craig, 2012).

Why Use Various Indicators of IER?

As illustrated above, each profile of survey participants interacts differently with IER indicators. In order to identify each profile of IER participants, screening should be performed using various indicators of IER. Therefore, employing diverse indicators is a necessary step that enables us to identify and classify distinct types of IER participants in the data screening process.

This reliance on multiple indicators is crucial given that the sample-wide sensitivity of IER indicators tends to be limited. As shown by Meade & Craig (2012), each indicator is sensitive to a specific response pattern. For example, while the long string index is very useful to identify IER participants providing overly repetitive strings of consecutive responses (e.g., responding “4” to all the items of a scale), the same indicator struggles to identify participants who provide completely random responses. As such, the selection of IER is not a process to be taken lightly, and researchers have thus been promoting the use of a “multiple hurdles approach” (Curran, 2016; Ward & Meade, 2023), which is also known as the sequential approach. As each

indicator seeks to capture distinct problematic response patterns, it is hoped that by sequentially screening the data using various indicators, it is possible to increase the overall sensitivity of the screening process by targeting various profiles of IER participants.

Limitations of IER Detection Approaches

The sequential approach (i.e., serial screening of the data using a series of indicators) is currently considered the gold standard in terms of IER detection (Curran, 2016; Ward & Meade, 2023). While the sequential approach might technically increase the sensitivity of IER by targeting more profiles of IER participants (Curran, 2016; Meade & Craig, 2012), numerous limitations surrounding the application of the approach remain unaddressed. Researchers implementing the sequential approach face questions regarding the selection of indicators (i.e., *what are the best indicators to identify IER participants*, *when should each indicator be used*), and their application (i.e., *which cut-offs should be used for each of these indicators*). In response to the scatteredness of the IER literature, an extensive review presenting the best practices in terms of IER detection and prevention was thus recently published (Ward & Meade, 2023). Although presenting best practices in terms of IER detection might improve the consistency of the IER indicators used by the researchers, such review fails to address the main limitations inherent to the sequential approach.

The main limitations of the sequential approach can be tied to its reliance on cutoffs. The sequential approach indeed requires the specification of cutoffs (i.e., the value/score beyond which one should be deemed IER) for each indicator (Barnes, 2018; Curran, 2016; Ward & Meade, 2023). Although multiple researchers have tried to find and specify generalizable cutoffs for each indicator (e.g., Desimone et al., 2015; Wang & Hau, 2024; Yentes, 2020), determining such cutoffs is not a simple task. The nature of the survey and the sample can impact not only the

scores yielded by each IER indicator but also their optimal cutoffs (Falk et al., 2025; Ilagan & Falk, 2023). The persistence of this gap in the literature poses a challenge because, in the absence of explicit guidelines, researchers must subjectively determine thresholds, which potentially impacts the quality of the screening process. Moreover, this gap entails that researchers end up using varying cutoffs for a same IER indicator. In other words, this implies that IER ends up being operationalized inconsistently across the literature. Not only does this complicate researchers' work, but such a gap also makes it impractical to replicate IER findings. Without a standardized approach to IER measurement, it remains impossible to portray the actual prevalence of IER across studies and the overall scientific literature. This highlights a need to introduce a novel and standardizable approach to IER detection. Furthermore, by establishing cutoffs, we limit the information provided by each indicator. For instance, when relying on Mahalanobis distance, the cutoff can be established at an alpha level of .05, .01, or .001. Participants with a *p*-value equal to or smaller than the established alpha level are deemed multivariate outliers and are subsequently removed from the analysis. Doing so, however, fails to recognize that participants with a very small Mahalanobis distance (thus a large and nonsignificant *p*-value) might technically be displaying signs of invariant (i.e., patterned) responding. Consequently, by imposing cutoffs, we lose potentially valuable information regarding the entire distribution of scores and certain types of IER participants. Similarly, when using individual variability, research recommends removing participants with high levels of variability (Marjanovic et al., 2015), thus failing to identify invariant IER response patterns (i.e., low levels of variability). In sum, the necessity to establish cutoffs for each indicator of IER minimizes the information that can be drawn from them.

Overall, these limitations highlight the need to move beyond the sequential approach and to introduce a novel IER detection approach that not only maximizes the information provided by each indicator but also tackles the limitations tied to the sequential approach's reliance on cutoffs.

Predicting Insufficient Effort Responding

Whilst IER indicators and detection approaches are useful to identify IER once it has already happened, the end goal should be to prevent IER from happening in the first place. This is especially true given the known pervasive impact of IER on research results (e.g., Arias et al., 2020; Huang, Liu, & Bowling, 2015; Kam, 2019). To do so, understanding the determinants driving the behaviour at a contextual and personal level is crucial to designing useful and meaningful interventions. In fact, past research has highlighted that various contextual (e.g., time of the semester, survey length; Bowling et al., 2021; Galesic & Bosnjak, 2009; Witt et al., 2011) and personal (e.g., personality, motivation; Ashley & Shaughnessy, 2021; DeSimone et al., 2018; Dunn et al., 2018) determinants can foster IER.

Contextual Determinants of IER

Contextual determinants refer to all the circumstantial, external, and environmental variables that may shape the context in which the participants complete the survey and ultimately impact how participants interact with said survey. Past research has investigated the role of various contextual determinants of IER, namely the role of survey length, study topics, and the time of the semester (i.e., when the participant completed the study within a given semester).

Survey Length

Results arising from a large body of research have already shown a positive relationship between IER and survey length (Bowling et al., 2021; Galesic & Bosnjak, 2009; Gibson &

Bowling, 2020; Ward et al., 2017). A study evaluating the prevalence of random responding (i.e., a type of IER behaviour) to the Minnesota Multiphasic Personality Inventory suggests that although random responding can be found from the start of a survey, it becomes even more prevalent toward the middle and end of the survey, especially in longer surveys (Berry et al., 1992). This suggests that IER might become more prevalent as the number of items to which participants are exposed increases. More research states that participants are more likely to drop from the study if the stated study length was longer, regardless of the actual study length (Galesic & Bosnjak, 2009). It was also found that participants in the longer survey condition were also more likely to provide longer strings of responses than participants assigned to the short survey condition (Galesic & Bosnjak, 2009; Herzog & Bachman, 1981). Results from a more recent study states that shorter surveys are associated with higher response and completion rates than longer surveys (Kost & Correa da Rosa, 2018).

In addition to evidence suggesting an effect of survey length on the prevalence of IER, results from various studies have shown that the prevalence of IER fluctuates as participants progress through the survey. More precisely, results from a study have revealed that IER increases at a within-person level as the participant progresses into a survey (Bowling et al., 2021). Evidence also suggests that the location of an item within a questionnaire (i.e., start vs. end) can influence the quality of the responses. Past research suggests that participants who responded to a given item at the beginning of a survey spent more time elaborating their response than participants who responded to the same item at the end of the survey (Galesic & Bosnjak, 2009); suggesting that the prevalence of IER increases as participants progress through a survey. Based on Berry and colleagues' (1992), there seems to be an interaction between the length of the survey and participants' engagement in IER at various instances throughout the survey. It,

however, remains unknown whether this effect persists after controlling personal determinants or for other contextual factors as study topics.

Study Topics

There is little literature that has investigated whether some study topics were tied with inherently more IER than others. Results from a recent study revealed that self-reported interest in the study topic was not significantly different across attentive and IER participants (Hasselhorn et al., 2025). However, this study solely considered the role of participants' self-reported interest on subsequent engagement in IER for a single topic. It remains unspecified whether the prevalence of IER differs in studies of various topics (i.e., arising from various research programs). Similarly, it is known that the reported prevalence of IER fluctuates across studies (Francavilla et al., 2019; Maniaci & Rogge, 2014; Meade & Craig, 2012). However, it remains unknown whether these differences are due to discrepancies in the IER detection approach or if these fluctuations are tied to differences in study topics or other contextual determinants. As such, there is a need to examine whether some topics are associated with a higher IER prevalence than others, while using a consistent IER detection approach.

Time of the Semester

Past research that analyzed the quality of the data collected through student recruitment pools found that participants completing their research participation credit at the start of the semester are less likely to engage in IER than participants taking part in research at the end of the semester (Aviv et al., 2002; Cheung, 2019; Witt et al., 2011). It is posited that participants' self-regulatory resources might be depleted once they reach the end of the semester, or that they are more motivated by the obtention of their course credits, which makes them more likely to engage in IER (de Calvo & Reich, 2007). It, however, remains unknown whether the time of the

semester is still associated with an increased prevalence of IER after controlling for other contextual determinants, as well as personal determinants of IER. More specifically, findings suggest that women and individuals with high levels of conscientiousness are also more likely to complete their participation earlier in the semester (Aviv et al., 2002; Witt et al., 2011), thus that they are less likely of engaging in IER. This highlights the relevance of also considering personal determinants of IER.

Personal Determinants of IER

Although contextual characteristics can either encourage or deter IER, some determinants of IER can be tied to personal characteristics. More precisely, past research has reported that participants' propensity to engage in IER differs across socio-demographic characteristics such as age, gender, and ethnicity (Braitman et al., 2025; Grau et al., 2019; Schneider et al., 2018). Previous research also suggests that participants' personality (Bowling et al., 2016; DeSimone et al., 2017, 2020; Dunn et al., 2018; Ward et al., 2017; Ward & Meade, 2023), emotional states (Ashley & Shaughnessy, 2021), and motivation (Cheung, 2019) can increase or decrease one's likelihood of engaging in IER.

Age

Studies have reported mixed results regarding the effect of age on one's propensity of engaging in IER. Although results from one study support that IER is more frequent in older participants (Schneider et al., 2018), another study failed found no significant effect of age after controlling for the effect of participants' personality, gender, age, education, and home country emancipation level (Grau et al., 2019). As such, more research is needed to assess whether the propensity to engage in IER differs across the lifespan while accounting for the effect of other personal determinants, such as ethnicity and gender.

Ethnicity

Although the effect of participants' ethnicity has not been widely assessed, a recent study suggests that the likelihood to engage in IER differed based on ethnic identity (Braitman et al., 2025). Another study, which has specifically examined cross-cultural differences in IER, found that the likelihood to engage in IER differed depending on participants' country of origin (Grau et al., 2019). While more research is needed to understand the origin of these differences, this suggests that cultural and social dynamics might be at play and impact one's propensity to engage in IER. Furthermore, it remains unknown whether this effect persists in studies of varying topics and when simultaneously accounting for other personal determinants such as gender, personality, and motivation.

Gender

Results arising from research posit that men are typically more prone to IER than women (Braitman et al., 2025; Cheung, 2019; Schneider et al., 2018). While the literature reports these gender differences, it is unlikely that one's gender causes IER. As such, more research is needed to understand whether gender differences persist while accounting for the effect of other personal determinants of IER, such as personality. Research indeed suggests that women tend to score higher than men on levels of agreeableness, neuroticism, conscientiousness, extraversion, and openness to experience (Weisberg et al., 2011). As such, it remains possible that the gender differences reported in the literature might arise from socialization or reflect gender-related personality differences.

Personality

The relationship between personality and IER has been the subject of investigation in the literature. The goal of these studies was to understand the role played by personality on

participants' engagement in IER and the overall quality of the data they provided. Results from past research suggest that self- and acquaintance-reported measures of extraversion, conscientiousness, agreeableness, and emotional stability were negatively related to IER (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017), and that IER is more strongly related to conscientiousness. These authors also state that IER was an enduring personal difference. However, evidence from research on this topic is mixed as it was found by Harms and colleagues (2017) that personality traits were not related to survey completion time.

The implications are serious if IER is driven by personality as it involves systematically removing participants based on individual differences, meaning that the representation in research of participants with such traits is limited. Therefore, more research is needed to determine if IER is mostly due to personality or if it is better predicted by motivation.

Motivation

IER is thought to be caused by a lack of motivation to read, understand, or respond to survey items (Huang et al., 2012). Although motivation is at the very core of its definition, the relationship between IER and motivation has yet to be thoroughly examined. Cheung (2019) provided preliminary evidence that a lack of motivation is associated with a higher prevalence of IER, but it remains unclear if motivation is truly the primary determinant of IER as doing so would require its examination while controlling for other personal and contextual determinants. More recently, Hasselhorn and colleagues (2025) have shown that a participant's average level of motivation (i.e., across time) was significantly associated with IER in ambulatory assessments, but that their motivation at an occasion-level (i.e., at each measurement) did not predict engagement in IER. While this provides us with initial evidence of the role of motivation, additional evidence is needed to support the claim that IER results from participants' motivation.

More specifically, these studies do not account for the role of other known personal or contextual determinants of IER. Moreover, these studies did not investigate the effect of motivation on IER in samples arising from diverse study topics and using a longitudinal prospective approach. Concretely, if motivation is the primary determinant of IER, this empowers researchers with interventions to increase participants' motivation, thus improving the quality of the data in their research. There is thus a need to conduct a comprehensive examination of the role played by participants' motivation on their engagement in IER.

Limitations of the Research on the Determinants of IER

Even if past studies have investigated the contextual (e.g., Aviv et al., 2002; Bowling et al., 2022; Cheung, 2019; Gibson & Bowling, 2020; Ward et al., 2017; Witt et al., 2011) and participants' determinants (e.g., Bowling et al., 2016; Camus, 2015; Dunn et al., 2018; Harms et al., 2017; Ward et al., 2017; Ward & Meade, 2023) of IER, multiple limitations remain. First, most of these studies have relied on a cross-sectional design, which prevents us from establishing causality. Second, these studies have investigated the individual role of contextual and personal determinants of IER without controlling for the effect of other known determinants of IER. Although doing so sheds light on possible antecedents tied to IER, it does not inform us about the most meaningful determinants of IER. Moreover, such methodological limitations imply, for instance, that is unknown whether study length is still related to IER after controlling for participants' motivation or accounting for the study's topic. Without relying on a prospective or longitudinal approach where personal and contextual determinants of IER are considered, it is impossible to establish whether IER constitutes an inter-individual differences (i.e., some individuals are simply more likely to engage in IER than others) as suggested by Bowling and colleagues (2016), or if it is instead better explained by motivational or contextual elements as

suggested by Maniaci & Rogge (2014). Nonetheless, our lack of understanding of the motives underpinning IER has serious implications, since it makes it impractical to design effective interventions targeting the roots of IER to prevent it from happening in the first place. It is why the present thesis sought to address this gap by jointly exploring the impact of contextual and personal determinants, but also by exploring the role of motivation. Doing so requires a robust and empirically valid theoretical framework. The present thesis hence explores the relationship between IER and motivation through the lens of Self-Determination theory (SDT).

Self-Determination Theory

Self-Determination Theory (SDT) is a leading theory in human motivation and is widely used across multiple domains (see Ryan & Deci, 2017, for an in-depth review of the theory). More specifically, SDT allows us to understand the reasons people engage and maintain certain behaviours. The notion of motivational quality is central to SDT. The theoretical framework suggests that individuals engage in various behaviours for different reasons depending on the degree of internalization of the behaviour (Ryan & Deci, 2017). It is stated that the emotional and behavioural benefits tied to the behaviour increase as a function of internalization. In other words, behaving in a way that is consistent with one's identity, values or interests, as opposed to external of internal pressure, is associated with greater behavioural commitment, perseverance and well-being. It is posited that there are three types of motivation (i.e., amotivation, extrinsic, and intrinsic), each differing in terms of motivational quality, or level of internalization. This internalization process can be further described along a continuum of six regulation types ranging from amotivation to intrinsic motivation (Ryan, 1995).

Starting from the least to the most internalized type of motivation, *amotivation* refers to behaviours enacted without intentionality or purpose; the behaviour is meaningless to the person

(e.g., participating in research although it feels useless). Next is *extrinsic motivation*. This type of motivation encompasses four types of regulation (i.e., external, introjected, identified and integrated). *External regulation* is the least internalized type of regulation within extrinsic motivation. Behaviours under such regulation are driven by external motives such as social pressure, rewards and/or punishments (e.g., participating in research to get a compensation). *Introjected regulation* refers to behaviours motivated by ego, shame, guilt, fear of disapproval, or that are executed in avoidance of such feelings (e.g., participating in research to avoid feeling guilty). *Identified regulation* refers to behaviours viewed as important to the person and performed for a specific purpose (e.g., participating in research because it is deemed important to the person). Finally, integrated *regulation* is the most internalized regulation and describes behaviours performed because they are in line with the person's values and/or goals (e.g., participating in research to achieve future goals). Lastly, there is *intrinsic motivation*. Intrinsic motivation is considered the optimal type of motivation since behaviours done with such motivation are performed simply for the satisfaction or pleasure of doing the behaviour itself (e.g., participating in research because it's enjoyable; Ryan & Deci, 2017).

Within SDT, it is stated that these regulations can be combined in various ways. Commonly, we distinguish either between self-determined and non-self-determined motivation, or between autonomous, controlled motivation and amotivation. The former relies on the creation of a self-determined motivation score by aggregating regulations that are self-motivated (i.e., intrinsic, integrated, and identified). A non-self-determined motivation score is obtained by aggregating the regulations that are external to the self (i.e., introjected, external, amotivation). In the latter approach, an autonomous motivation score is also computed by aggregating regulations that are self-motivated and driven by an internal perceived locus of causality (i.e.,

intrinsic, integrated, and identified) whereas controlled motivation encompasses regulation tied to internal and external pressures, in other words, to an external locus of causality (i.e., introjected and external), and amotivation, which is tied to an impersonal locus of causality, is treated individually. Although these regulations can be interpreted individually and provide researchers with granular insights regarding the role played by each type of regulation, issues with multicollinearity (Brunet et al., 2012) have brought researchers to regroup them either based on their level of autonomous (i.e., self-determined vs. non-self-determined motivation) or their locus of causality (i.e., autonomous vs. controlled motivation; Brunet et al., 2015). In the present thesis, the motivation types are grouped according of their perceived locus of causality and thus focus on autonomy (i.e., autonomy orientation), controlled (i.e., controlled orientation), and amotivation (i.e., impersonal orientation). This decision was guided by the fact that autonomous and controlled motivation are considered the two main forms of motivation within SDT (Vansteenkiste et al., 2005) and by previous research suggesting that amotivation might have a unique association with IER (Cheung, 2019). This distinction between autonomous, controlled, and amotivation has been shown to be relevant in various domains, including in sports (Spence & Oades, 2011), eating regulation (Guertin & Pelletier, 2023), and in education (Brunet et al., 2015; Ratelle et al., 2007; Vansteenkiste et al., 2005, 2009) where greater level of autonomous motivation has been tied to engagement (Vansteenkiste, Simons, et al., 2004) and persistence (Standage et al., 2003), two important elements when participants partake in surveys. Research has demonstrated that individuals who are guided by autonomous motivation tend to maintain their behaviours even when the difficulty of the task increases (Baxter & Pelletier, 2020; Green-Demers et al., 1997; Ryan & Deci, 2017). In contrast, controlled and amotivation have been tied to weaker behavioural engagement. Individuals guided by amotivation are more likely to become

disengaged if the difficulty of the task increases, if the external pressure is lifted, or if their engagement becomes an inconvenience (Lavergne & Pelletier, 2015). Taken together, this suggests that the type of motivation could shape how participants approach and interact with the survey in a research context. However, since motivation can play at various levels, it is not only important to consider its degree of autonomy but also its specificity.

Hierarchical Model of Intrinsic and Extrinsic Motivation

Within SDT, Vallerand's (1997) Hierarchical Model of Intrinsic and Extrinsic Motivation (HMIEM) states that motivation can be distinguished into three levels based on its degree of specificity (i.e., global, domain-specific, and situational). The least specific level, the *global level* of motivation, refers to general tendencies and one's personality. It reflects individual differences in how one tends to regulate and approach new behaviours through an autonomous, controlled, or amotivated orientation. Then, there is the *domain-specific level* of motivation. The domain-specific level of motivation is more specific than the global level. Instead of referring to general tendencies, the *domain-specific level* describes an individual's motivational orientations in a particular domain, such as education, sports, or research participation. It describes the motivational orientation that underpins one's actions within that specific domain. Lastly, the more specific level, the situational level of motivation, refers to moment-specific motivation within a specific domain, such as an individual's motivation orientation for participation in research as they start the survey. These three levels of motivation are hierarchical, from global (i.e., least specific) to situational (i.e., most specific).

According to the HMIEM, motivation at a higher level has a *top-down effect* on the motivation at a lower level. For example, motivation at a global level can influence contextual (e.g., research participation) motivation (Vallerand, 1997). In other words, someone who tends to

be globally autonomously motivated should tend to engage in research under a similar orientation (Lavigne & Vallerand, 2010). The HMIEM framework also supports that the motivational orientation at a lower level of specificity (e.g., domain-specific) can influence motivation at a higher level (e.g., global), exerting a *bottom-up effect*. Consequently, it is possible to posit that a respondent who tends to participate in research out of genuine interest (i.e., autonomous domain-specific motivation) should also experience higher levels of self-determined (autonomous) motivation generally in their life (i.e., autonomous global motivation).

Taken together, this highlights the relevance of investigating motivation at various levels of specificity. On one hand, examining the role of global motivation can help us better understand one's disposition towards any given task, irrespective of context. On the other hand, since global motivation acts as a trait, it is challenging to change one's level of global motivation. That is why considering motivation at a domain-specific level (i.e., for research participation) is also important as it represents a more malleable intervention target. In the context of IER, examining global and domain-specific motivation could help us understand whether IER arises from a general disposition or from a more specific motivational orientation.

Examining Insufficient Effort Responding Through the Lens of Self-Determination Theory

SDT offers a compelling framework for understanding the motives underlying participants' engagement in IER. In line with HMIEM postulates, previous research suggests that academic and research amotivation are positively associated with IER (i.e., completion time, semantic synonyms and antonyms, long string; Cheung, 2019). As well, it is supported that students' global autonomous motivation and global amotivation are respectively associated with a decreased and increased occurrence of IER. Therefore, participants who generally feel a sense of choice and meaning in their life were less likely to display signs of IER; whereas participants

who generally felt like they do not have a sense of control over their life were more likely to engage in IER. However, in a subsequent study, it was observed that academic amotivation completely explained the positive association between global amotivation and IER (Cheung, 2019). This suggests that engagement in IER was better explained by a more specific (i.e., domain-specific > global) level of motivation. Controlled motivation, regardless of the level (i.e., global and domain-specific), was not related to IER. Lastly, it was found that that amotivation at a global, domain-specific, and situational level (Cheung, 2019) are all positively associated with IER, but that more specific forms of motivation might be stronger correlates of IER.

Although Cheung's (2019) work provides us with insight on the role of motivation and hierarchical level of motivation on IER, more research is needed to better understand the relationship between motivation and IER to design targeted interventions. While Cheung (2019) considers the effect of the time of the semester, their study does not account for other personal determinants of IER and across studies of varying topics. As such, it is still unknown whether IER is mainly driven by motivation above and beyond other determinants such as personality traits. A thorough examination of the role of motivation, while accounting for these elements, is especially important given that motivation is defined as a lack of motivation to comply with survey instructions, interpret and provide reliable responses to items (Huang et al., 2012). Without doing so, it remains unknown whether IER truly arises from motivation, a lack of motivation (amotivation), or is better explained by other personal and/or contextual determinants. Examining this question at various levels of motivational specificity can help us understand whether IER arises from a general motivational disposition or, at a more specific level, from motivation towards research. Such distinctions have important implications since it is much easier to target domain-specific motivation than global motivation.

The Present Dissertation

Building on past research, the overarching goal of the present thesis was to improve the quality of research conducted through surveys by not only highlighting relevant intervention targets but also improving researchers' ability to identify IER participants. More specifically, this thesis had two main objectives. The first objective was to ***improve the detection of IER*** by providing evidence for the accuracy and validity of a novel IER detection tool, which circumvents the limitations tied to the sequential approach while maximizing the information provided by each indicator. The second objective was to ***highlight the contextual and personal determinants*** promoting participants' engagement in IER. More specifically, the goal was to shed light on the unique contribution of motivation by examining the motivational pathway underlying participants' engagement in IER, while accounting for known determinants of IER. To attain these objectives, a series of six studies, presented across three manuscripts, was conducted.

Structure of the Research Program

In *Manuscript 1,* the goal was to introduce the multidimensional IER detection approach (mIERda), a novel IER detection approach, across a broad set of conditions. Designed to address limitations surrounding the use of IER indicators, the criterion validity of the mIERda was assessed through a series of four simulation studies. In Study 1a, the classification accuracy of the mIERda was assessed and compared to the sequential approach using simulated attentive and IER data. In Study 1b, the classification accuracy of an optimized version of the mIERda was assessed under the same conditions. In Study 2, the classification accuracy of the mIERda is assessed using human attentive data and compared to the sequential approach. Finally, in

Study 3, the classification accuracy, sensitivity and specificity of the mIERda were assessed in a "multi-scale" setting (i.e., when two scales are being screened at once).

Building on *Manuscript 1*, *Manuscript 2* sought to provide further evidence of the validity of the mIERda. More precisely, in *Manuscript 2*, I rely on a prospective cross-sectional design to examine the construct validity of the mIERda using 18 datasets arising from five distinct research programs. More precisely, I assess the convergent (i.e., attention checks, self-reported items and survey completion time), concurrent (i.e., survey length, time of the semester), discriminant (i.e., controlled motivation at a global and research level), predictive (i.e., autonomous motivation and amotivation at a global and autonomous level), and known-groups (i.e., gender) validity of the mIERda. Overall, the objective was to examine the relationship of the mIERda within its broader nomological network by assessing its relationship with other indicators of IER and known correlates of IER.

After establishing the criterion and construct validity of the mIERda, I then examined the contextual and personal determinants of IER. In *Manuscript 3*, a prospective longitudinal design was used to investigate (a) the role of contextual determinants (i.e., survey length, study topic, and time of the semester) and (b) of participants' characteristics (i.e., personality, global motivation, research participation motivation) on IER. By relying on a prospective longitudinal approach, this allowed for the disaggregation of between- and within-person effects for contextual determinants. Overall, the objective of *Manuscript 3* was to explore the joint effect of contextual and participants' characteristics on odds of engaging in IER, and whether the effect of these determinants differ throughout the survey, while accounting for the variability associated with the participant. The specific objectives and the methodology underpinning each study will be presented in the following chapters of this thesis.

CHAPTER II

Sorting the Good from the mIERda – Introducing and Testing the Accuracy of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

Ariane J. Gauthier[1], Luc G. Pelletier[1] & Christopher Gravel[2]

Affiliation
[1]School of Psychology, University of Ottawa
136 Jean-Jacques-Lussier, Ottawa, Ontario, Canada K1N 9A8
[2]School of Epidemiology and Public Health, University of Ottawa
600 Peter Morand Crescent, Ottawa, Ontario, Canada K1G 5Z3

Author's Contributions Statement CRediT

Ariane J. Gauthier: Conceptualization, Methodology (lead), Software, Validation, Formal Analysis, Data Curation, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing (lead), Visualization
Luc G. Pelletier: Supervision, Writing – Review & Editing (supporting)
Christopher Gravel: Methodology (supporting), Writing – Review & Editing (supporting), Resources

Abstract

The scientific community is increasingly recognizing the relevance of addressing insufficient effort responding (IER). Nonetheless, the reliance of current IER detection approaches on cut-offs limits the implementation of certain indicators of IER and renders it difficult to identify IER, to estimate its prevalence, and to measure the impact of IER. The present article illustrates, across four studies, an innovative application of a fair-cut forest algorithm by introducing the *multidimensional Insufficient Effort Responding detection approach* (mIERda). Built on an unsupervised learning algorithm, the mIERda is an encompassing tool designed to streamline the IER screening process, address the methodological gaps associated with current detection approaches, and fit a maximum of survey designs. In *Study 1a*, a Monte Carlo simulation was conducted to assess the discriminative potential of various indicators of IER. The classification accuracy of the mIERda, in terms of IER detection, was also assessed and compared to the sequential approach across simulated data on three scales. In *Study 1b*, we assessed the classification accuracy of an optimized version of the mIERda. In *Study 2*, we assessed the classification accuracy of the mIERda in datasets containing real human data contaminated with simulated IER respondents. In *Study 3*, we simulated attentive and IER data to examine the classification accuracy, sensitivity, and specificity of the mIERda when multiple scales are screened at once. Overall, we found the mIERda to be more accurate than the sequential approach in simulated and human data. We also found the mIERda to have greater levels of balanced accuracy in longer scales and when the severity of IER was higher. Screening multiple scales at once did, however, improve the overall performance of the mIERda when working with scales containing fewer items. The impact of the prevalence rate on the classification accuracy of the mIERda depended on the type of simulated IER respondents

simulated. Overall, results provide evidence of the criterion validity of the mIERda and position it as an accurate IER detection approach.

Keywords: Insufficient effort responding; Careless responding; Content non-responsivity; Survey research methods.

"Sorting the Good from the mIERda" – Introducing and Testing the performance of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

Researchers from multiple fields and disciplines rely on surveys to find answers to their research questions. As these researchers collect data using surveys, the integrity of these data relies on the assumption that the responses provided by the respondents accurately portray their variables of interest. It is assumed that respondents are attentive and deploy the efforts required to properly read, interpret and respond to survey items. This belief is, however, being challenged by a growing body of research supporting that a significant number of respondents engage in insufficient effort responding (IER) when participating in survey research (Bowling et al., 2016; Huang et al., 2012; Huang, Liu, Bowling, et al., 2015). Also known as careless responding, IER is defined as a problematic response style characterized by low or little motivation to follow survey instructions, interpret item content, and provide accurate responses (Huang et al., 2012; Meade & Craig, 2012). As they engage in IER, respondents deviate from the core assumption that their responses result from an interaction between their latent trait and the items' characteristics. Instead, these respondents provide content-independent responses that are a representation of another unrelated underlying process.

Impact of IER

While early research suggested that IER had a relatively low prevalence in research samples (Johnson, 2005), it is currently estimated that IER respondents represent between 3% and 50% of research samples, in rates varying greatly between studies (Francavilla et al., 2019; Maniaci & Rogge, 2014; Meade & Craig, 2012). Unless they adequately address IER in their samples, researchers unknowingly end up conducting their analyses on partially spurious and potentially misleading data. If left unaddressed, this can affect the reliability of their results by

compromising the psychometric properties of their measures and bias the association between said measures (Arias et al., 2020; Goldammer et al., 2020; Huang et al., 2012; Meade & Craig, 2012). The common nature of IER is a source of concern, since evidence shows that even a low prevalence of IER can have a substantial impact on the integrity drawn from research data. IER had been shown to modify the factorial structure of scales (Arias et al., 2020; Huang, Bowling, et al., 2015; Huang et al., 2015; Kam, 2019; Woods, 2006), lead to inflated correlations (Huang et al., 2015), and deflated correlations (Kam, 2019), reduce scale reliability (Arias et al., 2020), reduce construct validity (Kam, 2019; Kam & Meyer, 2015), incur a loss of statistical power (Maniaci & Rogge, 2014; Oppenheimer et al., 2009), diminish effect sizes, and impact hypothesis test results (Maniaci & Rogge, 2014). By engaging in IER, these respondents unknowingly induce a form of measurement error which represents a considerable threat to data quality (DeSimone & Harms, 2018) impacting the reliability of survey research and its results. It is suspected that IER leads to the publication of spurious results, thus impacting the reproducibility of certain findings (Curran, 2016). While psychology finds itself in the midst of a replication crisis, with only 39% of findings and 47% of effect sizes being reproduced (Aarts et al., 2015); the survey-reliant field of social psychology finds itself particularly impacted with replication rates ranging between 20% and 45% (Aarts et al., 2015; Schimmack, 2020). Although it is unlikely that IER is the sole cause of the replication crisis, it remains a suspected culprit (Curran, 2016). It is ergo crucial to detect and address the presence of IER in survey data.

Identifying IER Respondents

Whilst the potential impact of IER is well documented, it is also known that IER often goes undetected, therefore unaddressed, by typical data cleaning procedures (Curran, 2016). As such, multiple indicators of IER (IERI; e.g., Mahalanobis distance, person-total correlation, long

string index, survey completion duration) have been introduced in the literature (Curran, 2016; Dupuis et al., 2018; Marjanovic et al., 2015; Meade & Craig, 2012; Schroeders et al., 2022) to facilitate the identification of these respondents.

Indicators of Insufficient Effort Responding

Various IERI have been introduced in the literature over the past years to facilitate the identification of IER respondents and limit its impact on data quality. For a full review of these IERI, refer to Curran (2016), Desimone et al. (2015), Huang et al. (2012); Huang, Liu, et al. (2015), and Ward and Meade (2023).

IER respondents identification is done by either (1) adding attention checks meant to measure respondents' attentiveness (e.g., instructed items and bogus items), (2) relying on paradata collected during data collection (e.g., completion time), or (3) by using a data-driven approach (e.g., Mahalanobis Distance; intra-individual variability; person-total correlation; long string index; psychometric synonyms and antonyms; even-odd consistency).

The first category of IERI embeds *instructed items* for which an expected response is provided to the participant (e.g., "*Respond strongly agree to this item*"; Gummer et al., 2018) and *bogus items* for which there is a sole correct response (e.g., "*I have never used a computer*"; Beach, 1989). These items are inconspicuously added amongst the other survey items and are designed to capture inattentive respondents. This approach is built on the assumption that IER respondents will be identified either by failing to properly respond to the item or by willingly disclosing their inattentiveness at the end of the survey. These approaches have been criticized since they disturb the flow of the survey, can annoy respondents, or prompt humoristic responses (Costa & McCrae, 2008; DeSimone et al., 2018) leading researchers to also rely on paradata and statistical methods.

The second category aims to identify IER respondents through the collection of paradata which relies on observing respondents' *completion time at a survey or page-level* (Bowling et al., 2023), *log data* (Kroehne & Goldhammer, 2018), *response latency* (Leiner, 2019), as well as *keystrokes and mouse clicks* (Kieslich & Henninger, 2017; Olson & Parkhurst, 2013). This approach has also been shown to be useful to identify bots (Buchanan & Scofield, 2018) which can be programmed to respond correctly to *instructed items.* The assumption underlying these indicators is that IER respondents interact differently with surveys than attentive respondents. However, these indicators cannot be obtained retroactively, which might limit their applicability in some instances (e.g., archival data, paper-and-pencil data collection). Statistical indexes are thus more suitable in such context.

The third category of IERI relies on different statistical indexes or data-driven indicators to facilitate the detection of IER respondents. Contrary to the two other categories of IERI, data-driven IERI rely on the direct analysis of the responses provided by the respondents rather than indirectly measuring carelessness through additional items or features (Karabatsos, 2003). This approach encompasses IERI that can be computed after data collection such as the *long string index* which aims to identify excessive string of repetitive responses (Johnson, 2005); *intra-individual variability* which screens for excessive intra-individual standard deviation (Marjanovic et al., 2015); *Mahalanobis distance* which screens for multivariate outliers (Meade & Craig, 2012); *person-total correlation* which is a measure of the within-person correlation between individual responses and sample-level mean responses for each item (Karabatsos, 2003); *psychometric synonyms* and *antonyms* which are measures of the within-person correlation between pairs of highly positively (synonyms) or negatively (antonyms) correlated items at a sample-level (Meade & Craig, 2012); *odd-even consistency* which is a measure of the

correlation between even and odd items (Huang et al., 2012; Meade & Craig, 2012); *response coherence* which is a measure of the predictability of the responses provided to items (Dupuis et al., 2015, 2018); and *response reliability* which captures the consistency between responses to similar items (Dupuis et al., 2015, 2018)

How do IER Respondents Differ in Terms of Response Patterns?

While all IER respondents exhibit signs of carelessness, not all of them manifest the same symptoms of IER. Not only can respondents differ in the onset and location of IER (i.e., IER from the beginning to the end vs becomes IER throughout the survey) within a given survey, but they also differ in terms of response patterns. Response patterns emerging from an IER process are often classified into two broad categories: (uniformly) random and non-(uniformly) random (Meade & Craig, 2012). A random response pattern refers to a response process where, for a given respondent, each Likert point has an approximately equal probability of being selected across items. This results in a response set with no distinguishable response pattern or trend, which could be modelled using a discrete uniform random distribution, although humans' responses cannot perfectly reproduce a fully random distribution (Figurska et al., 2008). The non-random category, in contrast, describes any content non-responsive response pattern where a distinguishable trend or tendency can be observed. Response sets under this category can be classified in two sub-categories: (a) invariant (e.g., responding "5" to all the items of a given scale), or (b) non-invariant where responses are positioned to both ends of the Likert scale (e.g., alternating between "1" and "7"), centred around the midpoint of the scale (alternating between "3", "4", and "5"), or patterned (e.g., responding "4-5-6-5-6", for example). The latter category could be modelled, respectively, by a bimodal distribution, normal distribution, or a Markov Chain.

Given that IER respondents differ in their response patterns, the sample-wide sensitivity of the most commonly used IERI (e.g., psychometric synonyms and antonyms, long string index, person-total correlation, etc.) is often limited. As demonstrated by Meade and Craig (2012), any given IERI is vastly more specific than sensitive in any given sample. This happens since each IERI is sensitive to a specific IER response pattern. For instance, while some IERI target excessive invariance (e.g., long string index), other target excessive variance (e.g., intra-individual variability) or inconsistency (e.g., person-total correlation). In other words, this means that using different IERI inherently leads to different operationalization of IER. As each IERI captures approximately distinct problematic response patterns, to increase the overall sensitivity of the screening process, researchers have been encouraged to use multiple IERI when performing IER screening. Such process is known as the sequential approach (Curran, 2016; Ward & Meade, 2023).

Sequential Approach

The sequential approach is an IER detection approach which relies on an iterative screening process of a selection of IERI. The sequential approach relies on a serial process where, in order to be deemed attentive, a respondent must “pass” each IERI. The implementation of the sequential approach relies on three major steps. To implement this approach, researchers must (1) select the IERI to be used, (2) establish a threshold (i.e., cutoff) for each IERI of what constitutes attentiveness vs IER, and (3) compute the IERI and screen the data using these cutoffs. Respondents whose score on any given IERI falls outside the prespecified cutoff, are deemed IER and are consequently excluded from the analytical dataset. This process is described in greater detail below.

Step 1 – Selection of the IERI. To implement the sequential approach, researchers must decide the combination of IERI to be computed and used for screening. Given the limited sensitivity of each IERI, this selection should seek to cover a maximum of response patterns (Meade & Craig, 2012). It is for this reason that various reviews were published to provide guidance to researchers interested in performing IER screening (Curran, 2016; Ward & Meade, 2023). Ward and Meade (2023) states that if IER screening is considered prior to data collection, researchers should consider adding instructed and bogus items, and to measure page-level response time. It is however worth noting that these IERI can solely be implemented if they are included to the survey prior to data collection. These IERI, consequently, cannot be implemented after data collection or in a simulation setting. Additionally, it is generally recommended for researchers to implement at least one IERI targeting invariance (e.g. long string index), another targeting multivariate outliers (i.e., Mahalanobis distance) and multiple IERI targeting consistency (e.g. psychometric synonyms and person-total correlation). For the needs of the present study, the sequential approach relies on the moderate screening level laid out by Ward & Meade (2023). The moderate screening approach states that the sequential approach should rely on one invariance indicator (i.e., long string index), one indicator of multivariate normality (i.e., Mahalanobis Distance), and two consistency indicators.

Step 2 – Selection of the cutoff for the IERI. After selecting the IERI to be implemented, researchers must determine a cutoff for each one. More precisely, the cutoff determines the threshold beyond which a respondent should be deemed IER. This is a necessary step since the sequential approach requires researchers to establish respondents' attentiveness for each IERI. Previous studies have consequently introduced and suggested rules of thumb to facilitate the implementation of these IERI (e.g., Barnes, 2018; Desimone et al., 2015; Wang &

Hau, 2024; Yentes, 2020). For instance, it is supported that a respondent should be considered IER if they consistently provide the same response for a number of instances equal to or greater than half of the scale length (Huang et al., 2012). The implementation of these cutoffs however remains somewhat subjective, and guidelines are unavailable for some IERI (e.g., response (Dupuis et al., 2015, 2018; coherence and response reliability).

Step 3 – Computation of the indicators. After selecting the cutoff, researchers must compute respondents scores on each IERI and remove any respondent who was flagged as IER by any given IERI. By doing so, it is hoped that researchers can improve the overall sensitivity of their screening process.

Limitations of the Sequential IER Detection Approach

The sequential approach requires researchers to select IERI and cutoffs to perform IER screening. However, most of past research pertaining to the identification of IER has been dedicated to the development of IERI and to the demonstration of their discriminative potential (Dupuis et al., 2018; Marjanovic et al., 2015; Meade & Craig, 2012). While doing so represents a first step to facilitate the detection of IER, researchers using these IERI must establish a cutoff value beyond which a respondent should be considered IER rather than attentive. Nonetheless, this step is often skipped in the literature which seriously limits their application. Furthermore, this reliance on cutoffs prevents the implementation of other promising IERI, such as response coherence and response reliability, for which they are no known or generalizable cutoffs. In previous research, these IERI were found to have a strong discriminative potential (Dupuis et al., 2015, 2018), nonetheless, the sequential approach's reliance on cutoffs limits their implementation.

While rules of thumb have been introduced to facilitate the application of some IERI (e.g., Barnes, 2018; Desimone et al., 2015; Wang & Hau, 2024; Yentes, 2020), the sequential approach's reliance on cutoffs limits the accuracy of the screening performed. Although relying on specific and pre-established may be intuitive, doing so fails to acknowledge that the optimal cutoff for each IERI may vary across scales and samples based on various factors (e.g., the number of items, factorial structure, correlation between items, the Likert scale of each scale; Falk et al., 2025; Ilagan & Falk, 2023). As such, implementing such fixed cutoff can impact the discriminative potential of each IERI, which ultimately can alter the quality of the screening performed. To address the complexities inherently tied to the use of the sequential approach, there is a need to move beyond cutoffs and to introduce an easily implementable data-driven ensemble approach. Reliance on a machine learning approach might represent a promising avenue.

IER Detection Driven By Machine Learning

The use of machine learning approaches has become increasingly common (Pugliese et al., 2021) and offers powerful opportunities to leverage the information provided by multiple indicators, without the need to establish specific cutoffs. Making these approaches especially promising in the context of IER detection.

In the recent past, IER detection approaches built on supervised (Schroeders et al., 2022) and unsupervised (Alfons & Welz, 2024) learning approaches have been introduced. Shroeder et al. (2022) applied supervised learning to IER detection using stochastic gradient boosted trees which resulted in poor performance in real-life settings due to the necessity of valid empirical training data. In contrast, the auto-encoder approach introduced by Alfons and Welz (2024) demonstrated better potential to address IER due to the unsupervised nature of the algorithm

used. However, the applicability of the approach to a broad audience remains limited due to the heavy computational nature of the algorithm. Machine learning approaches are also gaining traction in the field of bot detection (Falk et al., 2025; Ilagan & Falk, 2023). However, the applicability of such approach in an IER detection context is unknown. While bots necessarily provide content-independent responses to the entirety of items (Falk et al., 2025), humans engaging in IER might provide partially, inconsistent, or fully content-independent responses (Hong et al., 2020). Bots and humans thus differ in terms of their interactions with any given survey. Therefore, while we can expect an approach validated in an IER context to maintain a satisfactory level of performance for bots' detection, the performance is not guaranteed in the reverse direction. Contrarily to bots, intra-individual factors such as personality (Bowling et al., 2016; Grau et al., 2019), momentary motivation (Bowling et al., 2016; Hasselhorn et al., 2025), motivation towards research and global motivation (Cheung, 2019) can impact how, if, and when one engages in IER. As such, there remains a need for a performant and accessible machine learning approach to IER detection (i.e., low computational demand and limited data preparation); such approach has yet to be introduced in the literature.

In what follows, we build on this new stream of research and introduce the *multidimensional Insufficient Effort Responding detection approach* (mIERda) a novel IER detection approach driven by an unsupervised machine learning approach.

Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

Leveraging the advantages of a Fair-Cut Forest Algorithm (FCF), the mIERda proposes an innovative approach designed to be an easily implementable, interpretable, and standardizable IER detection tool for researchers working with survey data. Instead of interpreting IERI

individually or in a sequential manner, the mIERda proposes a multidimensional approach which iteratively screens the data using combinations of IERI using an adaptation of an FCF.

By doing so, not only does the mIERda minimizes researcher's degrees of freedom by relieving them from the burden of establishing cutoff at an IERI level, but doing so also maximizes the information provided by each IERI. Building on an FCF, the basic principle of the mIERda is to screen respondents using IERI, but to defer the interpretation of the scores for these IERI to the algorithm instead of vesting the pressure on researchers to establish cutoffs. By removing the need to establish cutoffs from the equation, this also opens the door to the inclusion of new IERI for which they are no known cutoffs (i.e., response coherence and reliability). Below we will discuss in greater details the algorithm underlying the mIERda and how it can be used to identify IER respondents.

Fair-Cut Forest Algorithm

The FCF originates from the Isolation Forest (IF; Liu et al., 2008), an unsupervised machine learning anomaly detection algorithm. The IF has been used to detect anomalies in various contexts such as cases of credit card fraud (Ounacer et al., 2018), health care data (Abirami et al., 2025), instances of cyber-attacks (Ahmed et al., 2019), anomalous test-takers (Bulut et al., 2024), and fraudulent survey data (Myöhänen, 2021). In contrast to other anomaly detection approaches, such as distance-based and clustering-based methods, IF does not profile normality in order to identify abnormal observations. Instead, IF relies on isolation trees (i.e., a type of decision tree) to build partial models by drawing subsamples from the data and by selecting indicators (i.e., features) uniformly at random. Such approach makes it more robust to swamping (i.e., labeling normal observations as anomalies) and masking (i.e., labeling anomalies as normal observations) than typical outlier detection approaches, especially in the absence of

abnormalities, in high dimensional problems and when some features are uninformative (Liu et al., 2008). This makes IF particularly suitable in a context of IER detection where the prevalence of IER can fluctuate greatly and where the informativeness of each IERI varies greatly depending on the response patterns displayed by IER respondents.

Over the years, multiple versions of IF were introduced to respond to specific needs. In fact, after past research has demonstrated that the initial version of the IF might struggle to identified “clusters” of anomalies (i.e., group of anomalous data points/respondents with similar characteristics) an adaptation of isolation trees, called fair-cut trees, were introduced to facilitate the detection of such instances (Cortes, 2021). In contrast with isolation trees which divides the data by applying uniformly random split, fair-cut trees work by applying splits which minimizes the pooled variance to facilitate the detection of clustered outliers (Cortes, 2021), making it particularly suitable to IER detection applications. An ensemble of fair-cut trees can be used to generate a fair-cut forest (FCF). Relying on an ensemble of fair-cut trees rather than individual trees can increase the robustness of the screening process and reduce error; it is why ensemble methods tend to be more accurate than approaches relying on a single classifier (Dietterich, 2000). Resulting anomaly scores range between 0 and 1; a score closer to 0 represents inlierness whereas a score closer to 1 represents a great likelihood of outlierness (Liu et al., 2008).

Identifying IER Respondents using the mIERda

Although iteratively screening the data using a FCF removes the need to set and establish cutoffs at an IERI level, which allows the algorithm to leverage all the information provided by each IERI, the need to establish a cutoff for the mIERda score remains. The mIERda is built on the assumption that we can distinguish between attentive and IER respondents by iteratively splitting the data based on the joint distribution of pairs of IERI. Building on the mechanism of

FCF, a data-driven cutoff can be established by finding the point which minimizes the variance in each resulting distribution. Doing so, the attentive and IER groups are defined by mIERda scores in rank order and split by a threshold which is chosen to minimize the between-group sum of squared errors. Such cutoff was selected as it represents a natural extension of the FCF splits applied within each tree.

Present Research

The main goal of the present research was to assess the classification accuracy and criterion validity of the mIERda under various conditions across four studies. To properly examine its criterion validity, the goal was not only to assess the classification accuracy of the mIERda, but to also compare it to the sequential approach. To do so we relied of a series of simulation studies.

In *Study 1a*, we sought to (a) examine the discriminative potential of the various IERI and (b) investigate whether the algorithm underpinning the mIERda improves the accuracy of the screening process (compared to the sequential approach). Data was simulated to assess the accuracy of the mIERda when screening data containing different types of non-invariant IER respondents across three psychometric scales of various lengths (items), dimensionality (factors), and response scale. To properly compare the two approaches, the mIERda and sequential approach were, in a first instance, computed using the same IERI. In *Study 1b*, we examined whether the classification accuracy of the mIERda improved by including the IERI with the greatest discriminative potential, as identified in *Study 1a*. We examined whether a more performant version of the mIERda could be obtained by including IERI that were not implemented (i.e., intra-individual variability) or could not otherwise be implemented using the sequential approach (i.e., response coherence and response reliability). In *Study 2*, we examined

the classification accuracy of this latest version of the mIERda in human attentive data contaminated by simulated IER respondents. Lastly, in *Study 3*, the classification accuracy of the mIERda at a survey-level was assessed by simulating responses across two scales.

Study 1a

Building on past research, *Study 1a* sought to examine the discriminative potential of the IERI with cutoffs, as prescribed by the sequential approach, and without cutoffs. The goal was also to assess the performance of the algorithm underpinning the mIERda under multiple conditions, across various scales, and to examine its criterion validity by comparing it to the current gold standard, that is, the sequential approach (Curran, 2016; Ward & Meade, 2023). We also sought to examine the discriminative potential of the IERI used in the sequential approach as well as of other well-known and high-performing IERI. More specifically, *Study 1a* sought to answer the following research questions (RQ):

RQ1 : What is the discriminative potential of the IERI when using the cutoffs prescribed by the sequential approach?

RQ2: What is the discriminative potential of the same IERI without implementing cutoffs?

RQ3: What is the discriminative potential of other known IERI?

RQ4: What is the classification accuracy of the sequential approach?

RQ5: What is the classification accuracy of the mIERda?

RQ6: Does using the mIERda, rather than the sequential approach, improves classification accuracy?

Methods

To address these questions, we performed a series of Monte Carlo simulations. A simulation approach was chosen as it is otherwise impossible to ensure the full attentiveness of

respondents in empirically collected data, and such approach limits sampling bias. We simulated data as to represent responses on three scales, namely the Experiences in Close Relationships (36 items; Brennan et al., 1998), the Big-Five Inventory (50 items; Goldberg, 1992), and the HEXACO Personality Inventory (240 items; Ashton et al., 2007). These scales were selected for their known and well-established multidimensional structure. For more detailed information regarding the structure of each scale, refer to Table 2.1. In order to examine the classification accuracy of the mIERda, data was generated while manipulating four parameters: the sample size, the prevalence rate, severity of IER, and IER response pattern.

Sample Size

The sample sizes (N) of the dataset simulated in this study were 500, 1000 and 2000. These values were chosen as to cover a common range of sample sizes observed in psychological research (Marszalek et al., 2011).

Prevalence Rate

The IER prevalence rates simulated in this study (π) were 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50%. These values were selected as they represent the commonly reported prevalence of IER in psychological survey research (Francavilla et al., 2019; Johnson, 2005; Meade & Craig, 2012).

Severity of IER

To simulate respondents that might have been attentive in the beginning of the survey, but ended up disengaging and turned IER, partially IER respondents were simulated. The severity of IER, more precisely the proportions (p) of survey items (k) as IER simulated in this study were P = 25%, 50%, 75%, and 100%. In the 25%, 50% and 75% conditions, the last $p*k$ attentive responses were replaced with IER responses. This was done as it is known that the

prevalence of IER increases as respondents progress through the survey (Bowling et al., 2021, 2022)

Attentive Responses

For each scale, attentive responses were generated from a graded 2-parameter IRT model using the *mirt* package (Chalmers, 2012). In order to ensure that the data was as realistic as possible, items parameters (i.e., discrimination and difficulty), and the inter-factor correlation matrix for each scale were estimated using publicly available empirical datasets obtained from an open source (openpsychometrics.org; Open-Source Psychometrics Project, 2019). Datasets from this website have been used in numerous scientific publications (e.g., Gonzalez et al., 2020; Goretzko & Bühner, 2020; Jacobucci et al., 2019) and are deemed to have a low level of IER since they were completed out of personal interest as respondents receive no compensation for their participation. However, in *Study 1a*, we simulated attentive responses from an IRT model to ensure that we have a greater control on the quality of the responses before testing the performance of the mIERda on real human data (*Study 2*). Respondents' latent traits (theta scores) were generated under a multivariate normal distribution, using the MASS package (Venables & Ripley, 2002).

IER Response Pattern

As to represent a broad spectrum of IER respondents, IER responses were generated under five random distributions and patterned probabilistic models (i.e., bimodal, moderately invariant pattern, highly invariant pattern, normal/midpoint, uniform). Although there is no consensus as it regards the best practices in terms of IER data simulation, these response distributions and patterns are generally accepted as valid non-invariant IER response patterns (Curran & Denison, 2019). Furthermore, generating multiple response patterns allows us to

understand how the mIERda performs in terms of detecting a specific IER response pattern. Assuming *k* denotes the number of items for each scale, the data generating process for these IER response patterns is described below.

Bimodal. In order to generate bimodal responses, a total of 10,000 samples were drawn from two normal distributions (*distribution 1:* $\mu = 1$, $\sigma = 1$; *distribution 2:* μ = number of response categories, $\sigma = 1$). To grant high probability to the responses at both ends of the response categories, both distributions were then merged and constrained within the lower and upper end of the scale's response categories to generate a bimodal distribution. The resulting distribution was discretized, and distribution probabilities were derived from observed frequency. Each respondent response vector of length *k* was then simulated based on the resulting probabilities.

Moderately Invariant. In order to generate the moderately invariant responses, a first value was drawn from a discrete random distribution (*min* = 1, *max* = number of response categories). Using the method described by Curran and Denison (2019), the following response had a 25% chance of decreasing by one scale point (e.g., from 2 to 1), a 50% chance of remaining the same, and a 25% chance of increasing by one scale point (from 2 to 3). This process was performed for *k*-1 items.

Highly Invariant. In order to generate the moderately invariant responses, a first value was drawn from a discrete random distribution (*min* = 1, *max* = number of response categories). Then, using the method described by Curran and Denison (2019), the following response had a 12.5% probability of decreasing by one scale point (e.g., from 2 to 1), a 75% chance of remaining the same, and a 12.5% chance of increasing by one scale point (from 2 to 3). This process was performed for *k*-1 items.

Midpoint. In order to generate midpoint responses, a vector of length *k* items was generated under a normal random distribution where μ is given by the median of response categories and σ = 1. The resulting distribution was constrained within the lower and upper end of the scale's response categories and discretized by rounding the resulting values.

Uniform. In order to generate the uniform responses, a vector of length *k,* was generated under a discrete uniformly random distribution bounded within the lower and upper end of the scale's Likert scale, and where each response choice has an equal probability of being selected.

Mixed. Since the goal of the present study is to test how the mIERda performs in a realistic setting, and as it is unlikely that all the IER respondents of a given sample would respond using the same response pattern, an additional *mixed* condition was generated by drawing an even number of rows of random responses generated from each distribution or patterned probabilistic model. As a result, a total of six conditions were considered for the IER response pattern parameter.

Plan of Analysis

For each scale, a total of 720 conditions were simulated: 3 (sample size) X 10 (prevalence rate) X 4 (severity of IER) X 6 (IER response pattern). Data was generated using a fully factorial design as to fully highlight the contribution and interaction between each factor. For each simulation condition, dataset was screened for IER using the sequential approach and the mIERda. In both instances, participants were screened using the same IERI.

The selection of the IERI to be implemented for the sequential approach and the mIERda was done in accordance with the moderate screening level laid out by Ward and Meade (2023). This screening level states that one invariance IERI (i.e., long string index), one IERI of multivariate normality (i.e., Mahalanobis Distance), two consistency IERI (i.e., person-total

correlation, and psychometric synonyms), completion time, and instructed and bogus items should be used. Considering the simulation nature of the present study, responses to instructed items, bogus items, and paradata were not available (i.e., respondents' page-level completion time is not available). Consequently, data were solely screened using the data-driven IERI recommended for this level of screening. More specifically, data were screened using long string index, psychometric synonyms[1], person-total correlation, and Mahalanobis distance. For the sequential approach, cutoffs were specified, in accordance with past research, for each IERI (refer to Table 2.2 for the cutoffs used for each indicator). Afterwards, IER screening was performed using the sequential approach and using the algorithm underlying the mIERda. In both instances, doing so returned a variable containing the attentiveness status for each "respondent" (0 = attentive, 1 = IER).

Respondents' mIERda score was obtained using the *isotree* package (version 0.6.1-1; Cortes, 2024). For the purpose of the present study, the hyperparameters of the FCF were set to 2 features, a subsample of 32, and 100 trees. A total of 100 replications were conducted for each simulated conditions in *R* (R Core Team, 2024).

Performance Measures

To estimate ability of the various IERI to differentiate attentive and IER respondents (i.e., discriminative potential), the area under the curve (AUC) was computed. In a first instance, the AUC of the IERI included in the sequential approach (i.e., long string index, Mahalanobis

---

[1] During data analysis, no pairs of psychometric synonyms ($r \geq .60$) could be retrieved in conditions where the prevalence of IER was equal or greater to 40%. To implement this indicator, pairs of highly correlated were thus identified in the original uncontaminated data, and used across all simulated datasets, instead of identifying these pairs individually for each simulated dataset.

distance, person-total correlation, and psychometric synonyms) was assessed, with and without a cutoff. Then, the AUC of other IERI (i.e., intra-individual variability, response coherence and response reliability) was also computed. An AUC closer to 0 and 1, is indicative of a strong discriminative potential. Conversely, an AUC closer to .50 is indicative of a poor discriminative potential.

Afterwards, the classification accuracy of the sequential approach and mIERda was assessed using three performance metrics, namely balanced accuracy, sensitivity, and specificity (refer to Table 2.3 for the formula used to compute each performance metrics). Although, accuracy is typically used to assess the proportion of correctly classified respondents amongst attentive and IER cases, its interpretation can be misleading in imbalanced datasets (Brodersen et al., 2010). Accuracy can be misleading because it may appear high even if the classification tool performs poorly in the minority class (i.e., smallest group). As such, the balanced accuracy was used, as it corrects for class imbalance by computing the average of correctly classified respondents across classes, which gives equal weight to both classes. A value closer to 1, on any of these measures, indicates a stronger classification performance, whereas a value close to 0 denotes a poor classification performance.

Results and Discussion

In the present study, we simulated data for three scales by following a 6 (IER response patterns) X 4 (Severity) X 10 (Prevalence rate) X 3 (Sample size) design. For the sake of conciseness, as results were similar across sample sizes, the results presented below corresponds to $N = 2000$, unless otherwise specified. In what follows, we firstly focus below on the interpretation of the AUC of the IERI. The balanced accuracy of the sequential approach and of the algorithm underlying the mIERda is then reported. Tabulated results for other performance

metrics (i.e., sensitivity and specificity) and for other sample sizes can be found in the Supplemental Material of *Manuscript 1*.

What is the discriminative potential of the IERI when implementing the cutoffs prescribed by the sequential approach?

When the IERI (i.e., Mahalanobis distance, long string index, psychometric synonyms and person-total correlation) were implemented using the cutoffs prescribed by the sequential approach, their discriminative potential was found to be fairly weak (Figure 2.1 – A). Although the AUC of Mahalanobis distance indicated a satisfactory discriminative potential across scales for bimodal (range: 0.84, 0.96) and uniform IER respondents (range: 0.73, 0.82) – the AUC suggested a low discriminative potential of Mahalanobis distance for other types of IER respondents. The AUC otherwise revealed a limited discriminative potential for the long string index (range: 0.50, 0.51), psychometric synonyms (range: 0.48, 0.65), and person-total correlation (range: 0.57, 0.62) towards each type of respondents. This suggests that the discriminative potential of the indicators, as currently implemented within the sequential approach, remains limited.

What is the discriminative potential of the same IERI without implementing cutoffs?

Implementing the same IERI, without cutoffs, resulted into a largely superior discriminative potential (Figure 2.1 – B). For instance, the AUC of the long string index for highly invariant IER respondents suggested a largely superior discriminative potential when no cutoff was used (range: 0.95, 1.00) than when a cutoff was applied (range: 0.50, 0.51). This effect was observed, without exceptions, across all conditions.

Amongst the IERI included in the sequential approach, highly invariant and moderately invariant respondents were better captured by the long string index, whereas bimodal and

uniform respondents were better identified by Mahalanobis distance. Midpoint and uniform respondents were better captured by person-total correlation, but solely in longer scales. Since other IERI did not display strong discriminative potential toward midpoint respondents, this suggests that IERI currently implemented within the sequential approach might fail to properly distinguish between attentive and midpoint IER respondents in shorter scales. While most IERI displayed a strong discriminative potential towards at least one type of IER response pattern, it was not the case for the psychometric synonyms indicator. This suggests that this indicator, with or without a cutoff, might not be as discriminative as other IERI.

What is the discriminative potential of other known IERI?

While not implemented within the sequential approach, the discriminative potential of additional IERI (i.e., intra-individual variability, response coherence, and response reliability) was investigated. As displayed in Figure 2.2, these IERI were found to have strong discriminative potential toward bimodal, midpoint and uniform respondents. While these IERI were not implemented within the sequential approach, these results suggest that their implementation could complement the other IERI and potentially result in a more accurate IER screening process. This is especially true given that currently implemented IERI have a limited discriminative potential towards midpoint and uniform IER respondents.

What is the classification accuracy of the sequential approach?

The balanced accuracy of the sequential approach fluctuated across scales, types of IER respondents (Figure 2.3 – A), and severity of IER (Figure 2.4 – A). Overall, the balanced accuracy appeared higher in Scale 3 (HEXACO; range: 29%, 93%), than in Scale 2 (BFI; range: 53%, 78%) or Scale 1 (ECR; range: 51%, 74%). This suggests that the sequential approach was more performant in longer scales. As it pertains to its classification accuracy

across types of IER respondents, the sequential approach tended to be more accurate when screening data solely containing bimodal IER respondents (range: 60%, 93%). In contrast, the classification accuracy tended to be weaker for highly invariant respondents (range: 29%, 58%) and moderately invariant respondents (range: 33%, 59%). This aligns with the limited discriminative potential of the various IERI toward these respondents when a cutoff was applied. Lastly, the balanced accuracy across conditions tended to be stronger when the severity was high (100% of the items, range: 50%, 93%), and decreased when the severity was weaker (e.g., 25% of the items, range: 49%, 71%), especially in shorter scales (i.e., Scale 1 – range: 49%, 53% & Scale 2 – range: 52%, 54%). Overall, results suggest that the sequential approach tend to be more accurate in longer scales, when IER respondents are solely bimodal, and when the severity of IER is high.

What is the classification accuracy of the mIERda ?

Similarly to what has been observed with the sequential approach, the balanced accuracy of the mIERda fluctuated across scales, type of IER respondents (Figure 2.3 – B), and severity (Figure 2.4 – B). The balanced accuracy of the mIERda tended to be higher in Scale 3 (range: 48%, 99%), than in Scale 2 (range: 46%, 89%) or Scale 1 (range: 34%, 86%). This suggests the mIERda might be more accurate in longer scales. In instances where the dataset contained solely one type of IER respondents, the mIERda tended to be more accurate when the scale was longer (Scale 3) and when the prevalence of IER was low (i.e., closer to 5%). As the prevalence increased, and the scales became shorter, the accuracy of the mIERda tended to decrease. This effect of the prevalence rate was not observed for the mixed respondents' condition. The accuracy tended to remain stable across prevalence rates when the data contained various types of IER respondents. Lastly, the mIERda otherwise performed best when the

severity of IER was high. The balanced accuracy tended to decrease as the severity decreased – this effect was, however, moderated by the scale. While the balanced accuracy of Scale 1 weakened substantially as the severity decreased, the balanced accuracy of Scale 3 remained more stable. Overall, results suggest that the mIERda tend to be more accurate in longer scale, when the dataset contains a mixture of IER respondents and when the severity is high. In contrast, the mIERda was found to be less accurate toward midpoint respondents and uniform respondents than towards other types of respondents.

Does using the mIERda, rather than the sequential approach, improves classification accuracy?

As displayed in Figure 2.3 and Figure 2.4, when screening the data using the same IERI, the mIERda tended to be more accurate than the sequential approach. This effect was especially noticeable in Scale 3 and when the data contained a mixture of IER respondents. Overall, this suggests that removing the needs for cutoffs and screening the data using mIERda resulted in an increased level of accuracy. More importantly, the mIERda was consistently more accurate than the sequential approach when screening data containing a mixture of IER respondents.

Summary

Taken together, results suggest that imposing cutoffs, as prescribed by the sequential approach, results in a significant loss of meaningful information and that an approach which allows for the implementation of these IERI without a cutoff is beneficial. Overall, these results support past research stating that each IERI is solely sensitive to certain IER response patterns (Craig & Meade, 2012) and highlight the relevance of relying on multiple IERI when performing IER screening.

Additionally, results suggest that the performance of the sequential approach and the mIERda depended on the severity, the prevalence, and the type of simulated IER respondents. Results highlight the impact of the type of simulated IER respondents on the performance of detection approaches (i.e., mIERda and sequential approach); with results demonstrating a differential performance across response patterns. Nonetheless, both detection approaches performed generally better when the proportion of IER (i.e., severity) increased and in scales containing more items. Results indeed support that the performance of the mIERda was weaker in Scale 1 (i.e., ECR; 36 items), but improved in Scale 2 (i.e., BFI; 50 items) and Scale 3 (i.e., HEXACO; 240 items), suggesting that the performance of the mIERda might improve as the number of items in a given scale increase. The effect of severity and scale length is unsurprising. In fact, a longer scale and a greater proportion of IER responses implies that the IERI has more information available to discern attentive from IER respondents. This also explains why the mIERda performed better when respondents were fully IER (high severity) than when they were partially IER.

The results suggest that the sequential approach might outperform the mIERda when the dataset contains a high prevalence of a single type of IER respondents (i.e., bimodal, midpoint, or uniform respondents). Conversely, the mIERda consistently outperformed the sequential approach across all conditions, irrespective of the scale, the prevalence rate, or the severity of IER, as long as the data included a mixture of IER respondents. Given that a mixture of IER respondents is expected in a real-life setting, this suggests that relying on an FCF algorithm rather than the sequential approach not only offers an avenue to standardize the IER screening process, but also to improve the quality of said screening. Results also revealed that the psychometric synonyms had limited discriminative potential, whether or not a cutoff was

applied. In contrast, other IERI, namely the intra-individual variability, response coherence and response reliability demonstrated significant discriminative potential. As such, it might be worth exploring whether adding these IERI to the mIERda, while excluding psychometric synonyms, would improve its classification accuracy.

Study 1b

After examining the discriminative potential of various IERI and whether using the mIERda could improve the accuracy of the IER screening process, we examined whether the classification accuracy of the mIERda could be further improved by optimizing the IERI used. As demonstrated in *Study 1a*, psychometric synonyms had limited discriminative potential across all types of respondents whether or not the IERI was implemented with a cutoff or not. In contrast, intra-individual variability, response coherence and response reliability demonstrated a strong discriminative potential towards bimodal, midpoint and uniform respondents. Building on *Study 1a*, the objectives of *Study 1b* were thus to assess the performance of optimized version of the mIERda by using the IERI with the most discriminative potential.

Method

Data was simulated using the methodology laid out in *Study 1a*. Rather than implementing the mIERda solely using the IERI prescribed by the sequential approach (i.e., long string index, Mahalanobis distance, person-total correlation, and psychometric synonyms), additional IERI were implemented, not only based on their performance in previous research, but also based on their performance in *Study 1a*. In the present study, we thus excluded psychometric synonyms due to its demonstrated limited discriminative potential. The mIERda was consequently implemented using the other IERI from the sequential approach (i.e., long string

index, Mahalanobis distance, and person-total correlation) and additional IERI (i.e., intra-individual variability, response coherence, and response reliability).

Results and Discussion

The balanced accuracy of the mIERda varied substantially across conditions and scales, ranging from 47% to over 100% (Figure 2.5). Results nonetheless followed similar trends to what had been observed in *Study 1a*, but the classification accuracy increased across conditions. While the classification accuracy remains relatively unchanged for mixed respondents (range: 72% – 96% vs 70% – 97%), optimizing the IERI implemented led to improvements in conditions where solely one type of IER respondents was simulated. For example, the balanced accuracy improved by 13% in Scale 1, at a 50% prevalence rate, when the data contained bimodal respondents. The improvement in classification accuracy was mostly noticeable in instances where the mIERda was performing more poorly in *Study 1a*. An increase in balanced accuracy was observed when the severity was low (i.e., when IER respondents provided IER responses to 25% of the items), the prevalence was high, in shorter scales (Scale 1 & 2), and when the data solely contained one type of IER respondent (i.e., bimodal, moderately invariant, midpoint, and uniform). This suggests that this optimized version of the mIERda could more accurately discern between these types of response patterns and simulated attentive respondents than the initial version of the mIERda.

Summary

Overall, results suggest that changing the IERI used within the mIERda had minimal impact of the classification accuracy when the data contained a mixture of IER respondents. Removing psychometric synonyms and adding intra-individual variability, response coherence, and response reliability however improved the accuracy of the mIERda when the data solely

contained one type of IER respondents. Since the characteristics of IER respondents are often unknown in an applied research context, this increased level of accuracy suggests this optimized version of the mIERda might be more robust than the initial version of the mIERda examined in *Study 1a*. While these results support the relevance of using this optimized version of the mIERda, its performance in an applied setting remains unknown. More specifically, while this study examined the classification accuracy of the mIERda in simulated attentive data, it is unknown how the mIERda performs in human attentive data contaminated by IER respondents. To demonstrate the potential of the mIERda in applied research setting where datasets contain various types of IER respondents, *Study 2* will evaluate the performance of the mIERda using real human data deliberately contaminated with a mixture of IER respondents.

Study 2

Study 1a and *Study 1b* established the performance of the mIERda across three scales using simulated attentive respondents. Its performance when working with "actual" human attentive data, however, remains unknown. Given that attentive empirical data is inherently "messier" than simulated data (Schneider et al., 2018), evaluating the performance of the mIERda in this context is essential to assess its practical usefulness. Therefore, the main objectives of *Study 2* were the following: (a) assess the performance of the mIERda using samples composed of empirical attentive respondents combined with simulated inattentive responses, (b) compare its performance to the sequential approach, and (c) compare its performance in a simulated vs empirical setting.

Methods

In order to test the performance of the optimized version of the mIERda in empirical data, mixed IER respondents were simulated using the methodology outlined in *Study 1a*. However,

instead of simulating attentive respondents, real human data were used. Specifically, attentive respondents were drawn with replacement from the dataset originally used to estimate the items parameters for *Scale 3* (HEXACO) in *Study 1a* and *1b*. The data was not screened for IER as to avoid introducing a collider effect. Since the performance of the mIERda remained stable across sample sizes in *Study 1a* and *1b*, to limit computational load, the sample size was fixed to $N = 500$ across conditions. The severity and prevalence of IER were manipulated using the methodology introduced in *Study 1a*. This process was performed for 100 iterations for all conditions.

Results and Discussion

The mIERda achieved higher classification accuracy (Figure 2.6 – B) when the severity of IER was higher. For instance, the classification accuracy of the mIERda was below 80% when the severity of IER was low (25% of the items). Contrastingly, the classification accuracy remained above 85% when the severity was high (100% of items). Globally, the balanced accuracy of the mIERda peaked when the prevalence was around 15% and tended to decrease slightly as the prevalence of IER increased. As displayed in Figure 2.6 – B*,* results also suggest that the mIERda was more accurate than the sequential approach regardless of the severity or the prevalence of IER. Both approaches also tended to be more accurate when the prevalence of IER was lower, and the severity was higher. Lastly, as demonstrated in Figure 2.6 – B, the mIERda was more accurate when the data contained simulated attentive data (vs. human attentive data; Figure 2.6 – A). This difference was especially noticeable when the prevalence rate was high and when the data contained partial IER respondents (severity = 25, 50, or 75%). Taken together, results suggest that the mIERda was more accurate, in terms of classifying respondents as IER or attentive, than the sequential approach.

Summary

We have examined the performance of the mIERda and the sequential approach across a wide range of prevalence and severity conditions. Overall, the results indicate that the mIERda outperformed the sequential approach in detecting IER across all tested conditions. Specifically, the mIERda displayed higher balanced accuracy compared to the sequential approach. However, a decline in classification accuracy was noted when comparing the mIERda's performance on real human data (*Study 2*) to that observed in simulated data for the same scale (*Study 1b – Scale 3*). While the exact cause of this finding is unknown, it is possible that the dataset from which attentive respondents were drawn was not completely exempt from IER; which could have introduced bias into our data. It is also possible that this effect is caused by the fact that real-world attentive responses tend to be more variable (Schneider et al., 2018), thus noisier, than simulated attentive data, which could explain the decreased accuracy of the algorithm in such condition. Nonetheless, results suggest that the mIERda was more accurate than the sequential approach across all conditions. Such results position the mIERda as a promising approach for IER detection in empirical settings.

Study 3

As demonstrated in *Study 1a* and *Study 1b,* the mIERda is more accurate when used to screen IER in scales containing more items. This however begs the question: Can the performance of the mIERda be improved by screening multiple scales at once? Furthermore, although evaluating the accuracy of the mIERda across individual scales is important to highlight which elements impacts its performance, it remains crucial to account for scenarios that more closely reflect real-world research applications. In practice, most researchers work with datasets containing multiple scales or measures. As such, the goal of *Study 3* was to test (a) how the

mIERda performs in a multiple-scale context and (b) to examine whether screening multiple scales at once can improve the classification accuracy of the mIERda.

Methods

To allow direct comparison with *Study 1b*, data was simulated as to represent responses to the *Experiences in Close Relationships* (Scale 1; 36 items) and the *Big Five Inventory* (Scale 1; 50 items). For an overview of these scales, refer to Table 2.1. Attentive and IER respondents were simulated using the same methodology as in *Study 1a*, excepted for the severity and sample size factor. Rather than specifying four fixed severity conditions, data were generated under two IER severity levels: (a) fully IER and (b) partially IER. In the fully IER condition, simulated IER respondents provided IER responses to all items. In the partially IER condition, the onset of IER for each respondent was drawn uniformly at random from a distribution ranging from 25% and 100% of the items. This modification was introduced to improve the generalizability of the findings, as it is unlikely that all IER respondents in a real-world setting would simultaneously start engaging in IER at the same point in a survey. Furthermore, due to the limited effect of sample size in *Study 1a* and *Study 1b*, and to limit computational load, datasets were simulated as to contain 500 respondents. For each scale, data was generated and merged in a single dataset as if each respondent responded to the two scales.

IER screening across multiple scales also bring a novel challenge as it regards the implementation of IERI. So far, the IERI were implemented for each scale. However, by using multiple scales, this opens the door to alternate forms of implementation. For instance, it would be possible to implement the IERI at a survey level (screening all the items at once, irrespectively of the scale). Similarly, it would also be possible to rely on a hybrid screening approach where some IERI are implemented at a scale level, while others IERI are implemented

at a survey level. While a hybrid approach would provide an uneven representation of each IERI, this could allow for the implementation of IERI that are currently only validated to be computed at a scale level (e.g., response coherence and reliability) while otherwise maintaining a survey-level screening. These various forms of screening can have an impact on the importance given to each IERI as well as on the sensitivity and specificity of the screening performed. For example, a survey-level screening should be less sensitive to partially IER respondents, while a scale-level screening should be more sensitive to such respondents.

In order to examine the effect of such implementation on the mIERda's classification performance, we examined and compared three levels of screening: at a scale-, survey-, and hybrid-level. As such, once the data generated, IERI (i.e., long string, intra-individual variability, Mahalanobis distance, person-total correlation, response coherence and response reliability) were implemented according to each level of screening. At a scale level, each IERI were computed for each scale. At a survey level, each IERI was computed across the entire survey, but only the highest values of response coherence and reliability across both scales were retained (to ensure an even representation of each IERI). In contrast, the hybrid screening level preserved the response coherence and reliability for each scale. For each level of screening, the resulting IERI were then processed simultaneously by the FCF algorithm. This was done to assess whether the mIERda performs better with IERI computed at a survey-, hybrid-, or scale-level.

Results and Discussion

The performance of the mIERda across prevalence, severity rate, and screening level was assessed using balanced accuracy (Figure 2.7), sensitivity and specificity (Figure 2.8). As displayed in Figure 2.7, regardless of the screening-level, the mIERda was more accurate when IER respondents were fully IER (range : 86% – 98%), than when they were partially IER (range:

76% – 85%). As shown in Figure 2.8, this can be tied to the fact that the mIERda tended to be more sensitive when the data contained fully IER respondents.

In the fully IER condition, the survey-level screening offered a slightly higher level of balanced accuracy than other level of screenings (i.e., scale- and hybrid-level). As shown in Figure 2.8, this effect is due to the increased specificity tied to the survey-level screening.

In the partially IER condition, all screening levels offered a similar level of balanced accuracy. However, whereas the hybrid screening level offered more sensitivity, the survey level screening offered more specificity. The hybrid screening level however offered the best level of accuracy and specificity when the prevalence rate was low (5% prevalence).

When compared to the results from *Study 1b,* we can notice that the performance of the mIERda tended to improve when screening multiple scales at once. In *Study 1b*, the balanced accuracy of the mIERda in the fully IER condition ranged between 76% and 81% in Scale 1 and between 79% and 90% in Scale 2. Under the same conditions, the balanced accuracy ranged between 86% and 97%, depending on the level of screening. In the partially IER condition, the balanced accuracy estimate ranged between 76% and 85% depending on the level of screening. In contrast, the balanced accuracy ranged between 71% and 81% (Scale 1) and between 76% and 83% (Scale 2) in *Study 1b* when pooling the severity levels. As such, results suggest that screening multiple scales at once does seem to increase the performance of the mIERda, or at least, does not impair it.

Summary

Overall results suggest that the mIERda retains a satisfactory level of performance when screening two scales at once. While more research is needed to examine this effect when more scales are being screened at once, results suggest that the optimal level of screening depends on

the objective of the researcher. Someone who wants to maximize the specificity of its screening, would benefit from a survey-level screening. In contrast, a researcher who want to maximize the sensitivity of their screening might benefit from relying on a hybrid screening approach especially when the data contains respondents with a mixed onset of IER. Since response coherence and reliability were overrepresented in the hybrid condition (i.e., they were implemented at a scale level whereas other IERI were implemented at a survey level), this suggests that these IERI might be more sensitive than other indicators, at least when the data contains a mixture of IER respondents. Nonetheless, while small variations were observed across levels of screening, the performance of each level remained, all things considered, relatively stable. This suggests that using the IERI and interpreting their information using the mIERda matters more than how each IERI is specifically implemented. Lastly, results suggest that shorter scales might benefit from being screening alongside other scales and that doing so can increase the accuracy of the screening and its sensitivity, and at worst, offer a similar performance.

General Discussion

The overall objective of the present research was to introduce the mIERda and provide preliminary evidence of its performance as it pertains to the detection of IER respondents under various conditions. To achieve this goal, the present research examined across four studies the performance of the mIERda, under various conditions (i.e., prevalence, sample size, severity & response pattern), in simulated and in empirical data. In a first instance (*Study 1a*), data was simulated to assess the discriminative potential of various IERI and assess performance of the mIERda in three scales of varying length (i.e., number of items), and factorial structure (i.e., number of factors) using the IERI suggested by the sequential approach. After demonstrating the advantage of using the mIERda over the sequential approach (*Study 1a*), an optimized version of

the mIERda was implemented in *Study 1b* using the most discriminative IERI. After examining how this optimized version of the mIERda performs in simulated data, we assessed in *Study 2*, its performance in empirical data and compared it to the sequential approach. Lastly, after noticing that the mIERda offered a weaker performance in shorter scales, another simulation study (*Study 3*), was conducted to assess whether the mIERda recovered a certain level of performance in a multiple-scale setting, that is when screening multiple shorter scales at once.

Main Findings

Discriminant Potential of the IERI

In *Study 1a*, we examined the discriminative potential of the IERI implemented within the sequential approach (i.e., with cutoffs) and without cutoffs. We also sought to examine how their discriminative potential varies depending on the types of IER response patterns. Doing so, we noticed that the discriminative potential of the IERI did fluctuate depending on the simulated response pattern. This corroborates past research stating that the sample-wide sensitivity of the IERI is limited (Craig & Meade, 2012). Furthermore, when examining the discriminative potential of the IERI, it was noticed that the psychometric synonyms displayed a limited discriminative potential toward each simulated type of IER respondents, whether or not a cutoff was implemented. In contrast, other IERI, namely the long string index, person-total correlation, intra-individual variability, response coherence, and response reliability demonstrated significant discriminative potential towards at least one type of IER respondents when no cutoff was used. This not only highlights the varying sensitivity of each IERI, but also the relevance of using multiple IERI in order to better capture IER.

The Pervasive Impact of Implementing Cutoffs

In *Study 1a*, results revealed that imposing cutoffs did negatively impact the discriminative potential of the IERI. This suggests that the implementation of cutoffs, as prescribed by the sequential approach, impairs the information provided by the IERI. Although we failed to compare how imposing various cutoffs can impact the discriminative potential of the IERI, we relied on commonly accepted and recommended cutoffs. Our goal was not to highlight optimal cutoffs, as they would not necessarily be generalizable (Falk et al., 2025; Ilagan & Falk, 2023), but rather demonstrate the potential impact using such cutoffs can have on the quality of the IER screening and highlight the potential limitations tied to the sequential approach.

mIERda vs. Sequential Approach

In line with the idea that the discriminative potential of the IERI was severely and negatively impacted by the implementation of cutoffs, the mIERda outperformed the sequential approach. This effect was observed in *Study 1a* when the sequential approach and the mIERda were implemented using the same IERI. This suggests that the FCF algorithm underpinning the mIERda led to an improvement in classification accuracy when performing IER screening. A similar effect was also observed in *Study 2* where the optimized version of the mIERda was found to outperform the sequential approach when screening IER in human attentive data. Taken together, these results provide evidence that the mIERda might represent a more accurate and performant approach to IER screening than the current gold standard (i.e., sequential approach).

The effect of Scale Length

In *Study 1a* and *Study 1b*, it was found that the mIERda performed best when used for IER screening in Scale 3 (240 items) and performed worst when screening in Scale 1 (36 items). While more research is needed to examine this effect, results provide preliminary evidence that

the mIERda might be more accurate when screening scales containing more items. This finding was later corroborated by results from *Study 3*, which demonstrated that simultaneously performing IER screening on two scale could improve the classification accuracy of the mIERda.

The effect of Types of IER respondents

In *Study 1a* and *Study 1b*, it was also found that the type of IER respondents simulated impacted the performance of the mIERda. More specifically, it was found that the mIERda performed best when it is used for IER screening in a dataset containing a mixture of IER respondents. Such result was expected considering the nature of the FCF which underlies the mIERda. The algorithm is designed to detect clusters of outliers (Cortes, 2021). As such, it was expected that the mIERda would perform best in such condition. This provides a positive outlook of the performance of the mIERda in a real-life setting, where researchers are faced with a mixture of IER response patterns (Hasselhorn et al., 2025; Meade & Craig, 2012).

The effect of Severity of IER

Across studies, it was found that the mIERda performed best when IER respondents were fully IER than when they were partially IER. Such finding is not unsurprising considering that as the proportion of inattentive/IER responses provided decreases for a given respondent, the difference between said respondent and fully attentive respondents should also decrease. Technically, a participant who engaged in IER for the last quarter of the survey provided, ratio-wise, more attentive than IER responses. As such, they could be considered “mostly” attentive. In such context, it is only logical that the performance of the mIERda decreases as it becomes harder for the algorithm to identify those cases.

The effect of Prevalence Rate

It was found that the balanced accuracy of the mIERda tends to decrease as the prevalence rate increases for most individual IER response patterns (e.g., midpoint). The effect of prevalence rate was however minimal for mixed IER respondents, as the balanced accuracy remained more stable across prevalence rates in such instances. Globally, this suggests that the classification accuracy of the mIERda remains consistent as long as the data contains a mixture of IER respondents. Such results offer a positive outlook of the mIERda in a real-life setting where a mixture of IER respondents is to be expected.

Simulated vs. Human Attentive Data

Results from *Study 2* suggest that the classification accuracy of the mIERda was weaker in empirical than in simulated data, especially as the prevalence of IER increased. Such results are not unsurprising and should not be interpreted as a strict limitation of the mIERda. Although the real human data was deemed to have a low risk of IER, the presence of IER remains possible. Nonetheless, it was found that the mIERda was more accurate than the sequential approach. This supports the idea that the mIERda performs better than the sequential approach in real human data.

Multiple-Scale Screening

In *Study 3*, results allow us to conclude that the mIERda retains a satisfactory level of performance when screening more than once scale at once. While the mIERda does not perform as well with shorter scale than with longer scales, screening two scales at once seemed to improve, or at least maintain, the classification accuracy of the mIERda. This suggests that while the mIERda does not perform optimally when individually screening shorter scales, the performance not only improves when screening longer scales, but can also improve when

performing IER screening across multiple scales. This aligns with the idea that measurement error decreases as the amount of information (i.e. number of items) increases (Cronbach, 1951; Nimon et al., 2012).

Screening Level

As it regards the effect of the level of screening, although the difference in classification accuracy was modest across approaches, results suggest that the approach to be selected should depend on the objectives of the researcher.

While computing IERI at a survey-level improved the specificity in the fully IER condition, the hybrid screening offered a better performance in terms of sensitivity when the data was fully and partially IER, especially as the prevalence of IER increased. Results suggest that researchers who may be worried that IER respondents may be impacting their research results and thus want to remove as many IER respondents as possible should aim to maximize the sensitivity of their screening by relying on a hybrid screening approach. In contrast, researchers worried to remove too many attentive respondents or working with a limited sample could maximize the specificity of their screening by relying on a survey-level screening.

Although, it is worth noting that these recommendations are non-exhaustive and that more research is needed to assess whether they hold true in other contexts. For example, it is likely that a survey-level screening may lack sensitivity when screening a large number of scales at once. Furthermore, it is also possible, that the hybrid-level screening leads to an over-representation of the response coherence and reliability indicators. Since they are computed at a scale-level, their representation is expected to increase as the number of scales in the survey increases. More research thus is needed to assess whether these factors can impair the performance of the mIERda.

Rethinking our Approach to IER Detection

The present research offers a compelling demonstration of the potential and performance of the mIERda across a broad number of conditions and positions the mIERda as an accurate IER detection tool. Results from the present research positions the mIERda as a potential substitute to the sequential approach as it pertains to the detection of IER. Overall, results suggest that while the sequential approach might be more accurate to detect specific types of IER respondents, the mIERda outperformed the former in the vast majority of instances when a dataset contained a variety of IER respondents to be identified. While the classification accuracy of the sequential approach was better under some conditions for certain types of IER respondents, the classification accuracy of the mIERda was generally found to be more consistent and predictable across prevalence rates, severity of IER, and when screening a mixture of IER response patterns.

While more research is needed to assess the performance of the mIERda across more conditions (e.g., number of items, Likert scales, number of scales), the present research serves as a demonstration of the potential of the mIERda. As supported by the results from *Study 1a, Study 1b* and *Study 2*, the mIERda was generally more accurate than the sequential approach. These preliminary results position the mIERda as an alternative option for researchers interested in detecting IER in their research sample.

Although the present research compares the sequential approach and the mIERda, it is however worth noting that the two approaches are not inherently in opposition. In fact, both approaches recognize the relevance of relying on multiple IERI in order to properly capture such respondents. Where both approaches diverge, is in terms of how they leverage the information of these IERI. The mIERda provides a way to standardize the IER screening process and leverage commonly used IERI without falling victim to their limitations such as cutoff specification. As

demonstrated in the present research, although IERI play an important role in IER detection, how the information they provide is being used might be even more important.

Limitations & Future Research

While the present research provides initial evidence of the performance of the mIERda, limitations pertaining to the study design and the optimization of the algorithm remain to be addressed.

Study Design

Although the present research assessed the feasibility of the mIERda and compared its classification accuracy to the current gold standard across a broad range of conditions, some elements remain to be investigated. Firstly, it is worth noting that the present research focuses on datasets where the sample size ranged between 500 and 2000 respondents. Even if results suggest that the performance remain consistent across this range, it remains unknown how the mIERda performs in smaller and larger samples. Beyond the limitations tied to the sample size, we also examined how the mIERda performed for three multidimensional psychometric scales without manipulating the number of items, item parameters (i.e., difficulty and discrimination), the factorial structure, or the number of response categories. As such, more research is needed to examine how the mIERda performs when these factors are manipulated. More research is also needed to understand whether, if so how, the presence of negatively worded items and inter-factorial correlation impact the performance of the mIERda. Future research would benefit from assessing how shifting and manipulating these parameters can impact the performance of the mIERda.

Furthermore, the scope of this paper was limited to non-invariant types of IER respondents. Based on the performance of the mIERda as it pertains to the detection of response

pattern with a certain level of invariance (i.e., moderately invariant and highly invariant) and considering that the mIERda relies on IERI targeting response variability (i.e., long string, intra-individual standard deviation, Mahalanobis distance), we can suppose that mIERda would capture fully invariant IER respondents, however its performance under such condition remains unknown.

Additionally, the performance of the mIERda was solely tested for IER detection at a scale-level (*Study 1a*, *1b* and *2*), then at a survey-level (*Study 3*) which contained solely two scales. While results suggest that the mIERda maintained a relatively stable level of performance when screening multiple scales at once irrespectively of the level of screening (i.e., scale-, hybrid-, and survey-level screening); this might not hold true in every instances. It is likely that in a survey containing more scales, computing indicators at a survey-level would substantially decrease the sensitivity of the mIERda. Furthermore, the performance of the mIERda in a multiple scale was solely assessed using simulated data. As such, more research is needed to examine this effect using surveys of differing length, other psychometric scales, and in empirical data.

Algorithm Optimization

The goal of the present research being to pilot the feasibility of implementing an IER detection approach relying on an FCF, we did not assess the impact of manipulating the algorithm's hyperparameters (i.e., sample size, number of trees, number of features) on the mIERda's performance. As such, it is likely that the performance of the algorithm can be further optimized, and future research would benefit not only from examining how manipulating these hyperparameters impacts the accuracy of the approach, but also from exploring whether other scales, samples and survey characteristics interact with the algorithm's hyperparameters. Doing

so would provide meaningful insight on whether the algorithm's hyperparameters can be generalized to various samples or if they should be adjusted based on these factors.

More research is also needed to investigate whether using other cutoffs computation approaches (e.g., Bayesian estimation or classification) could improve the classification performance of the mIERda. As demonstrated across studies, the performance of the mIERda tend to decrease as the severity decreases, as such, it is likely that the cutoff computation approach has a role to play on that front. As such, more research is required to examine whether the cutoff computation approach can be improved.

Beyond the need to investigate how hyperparameters and the cutoff can be tweaked to improve the mIERda, more research is also needed to establish whether adding other IERI, or removing some, can improve the performance of the mIERda. While the IERI currently included in the mIERda were selected for (a) their demonstrated discriminative potential and (b) their applicability of a broad number of research designs; future research would benefit from thoroughly investigating the contribution, and limitations, of each IERI. Accordingly, the list of IERI used in the present study is non-definitive and warrants further investigation.

Conclusion

Overall, results from the present study provide preliminary, yet compelling, evidence of the performance of the mIERda. Results from the present study positions the mIERda as a performant IER detection approach and provide evidence of its criterion validity. While room for improvement remains, the mIERda was found to be more accurate than the sequential approach across most conditions. This further supports the potential of approaches built on machine learning, such as the mIERda, as it pertains to the detection of IER.

References

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Abirami, M., Pavithra, S., & Stephen, K. R. (2025, January 15). Anomaly Detection In Health Care Data For Early Detection of Thyroid Diseases using Isolation Forest Model and Ridge Regression. *Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024)*. https://doi.org/http://dx.doi.org/10.2139/ssrn.5083332

Ahmed, S., Lee, Y., Hyun, S.-H., & Koo, I. (2019). Unsupervised Machine Learning-Based Detection of Covert Data Integrity Assault in Smart Grid Networks Utilizing Isolation Forest. *IEEE Transactions on Information Forensics and Security*, *14*(10), 2765–2777. https://doi.org/10.1109/TIFS.2019.2902822

Alfons, A., & Welz, M. (2024). Open science perspectives on machine learning for the identification of careless responding: A new hope or phantom menace? *Social and Personality Psychology Compass*, *18*(2). https://doi.org/10.1111/spc3.12941

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*(6), 2489–2505. https://doi.org/10.3758/s13428-020-01401-8

Arias, V. B., Ponce, F. P., & Martínez-Molina, A. (2023). How a Few Inconsistent Respondents Can Confound the Structure of Personality Survey Data: An Example with the Core-Self Evaluations Scale. *European Journal of Psychological Assessment*, *39*(6), 400–407. https://doi.org/10.1027/1015-5759/A000719

Ashton, M. C., Lee, K., & Goldberg, L. R. (2007). The IPIP–HEXACO scales: An alternative, public-domain measure of the personality constructs in the HEXACO model. *Personality and Individual Differences*, *42*(8), 1515–1526. https://doi.org/https://doi.org/10.1016/j.paid.2006.10.027

Barnes, T. (2018). *Determining Cutoffs for the Psychometric Synonyms Analysis To Detect IER* [Doctoral Thesis, Wright State University.]. Core Scholar. https://corescholar.libraries.wright.edu/etd_all/2013

Beach, D. A. (1989). Identifying the Random Responder. *The Journal of Psychology*, *123*(1), 101–103. https://doi.org/10.1080/00223980.1989.10542966

Bowling, N. A., Gibson, A. M., & Desimone, J. A. (2022). Stop With the Questions Already! Does Data Quality Suffer for Scales Positioned Near the End of a Lengthy Questionnaire? *Journal of Business and Psychology*, *37*, 1099–1116. https://doi.org/10.1007/s10869-021-09787-8

Bowling, N. A., Gibson, A. M., Houpt, J. W., & Brower, C. K. (2021). Will the Questions Ever End? Person-Level Increases in Careless Responding During Questionnaire Completion. *Organizational Research Methods*, *24*(4), 718–738. https://doi.org/10.1177/1094428120947794

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who Cares and Who Is Careless? Insufficient Effort Responding as a Reflection of

Respondent Personality. *Journal of Personality and Social Psychology*, *111*(2), 218–229. http://dx.doi.org/10.1037/pspp0000085.supp

Bowling, N. A., Huang, J. L., Brower, C. K., & Bragg, C. B. (2023). The Quick and the Careless: The Construct Validity of Page Time as a Measure of Insufficient Effort Responding to Surveys. *Organizational Research Methods*, *26*(2), 323–352. https://doi.org/10.1177/10944281211056520

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). The Guilford Press.

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010, August 23). The Balanced Accuracy and Its Posterior Distribution. *2010 20th International Conference on Pattern Recognition*, 3121–3124. https://doi.org/10.1109/ICPR.2010.764

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*, *50*(6), 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Bulut, O., Gorgun, G., & He, S. (2024). Unsupervised Anomaly Detection in Sequential Process Data. *Zeitschrift Für Psychologie*, *232*(2), 74–94. https://doi.org/10.1027/2151-2604/a000558

Chalmers, R. P. (2012). mirt : A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Cheung, S. (2019). *Understanding Insufficient Effort Responding from a Self-Determination Theory Perspective* [Doctoral Thesis, University of Ottawa]. uO Research. https://ruor.uottawa.ca/items/5db7e905-2c10-42a6-a335-6bf73341c438

Cortes, D. (2021). *Revisiting randomized choices in isolation forests* [Unpublished manuscript]. http://arxiv.org/abs/2110.13402

Cortes, D. (2024). *isotree : Isolation-Based Outlier Detection* (version 0.6.1-1). https://github.com/david-cortes/isotree

Costa, P. T., Jr., & McCrae, R. R. (2008). *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing* (D. Saklofske, Ed.; Vol. 2). SAGE Publications. https://doi.org/10.4135/9781849200479

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Curran, P. G., & Denison, A. J. (2019). *Creating Carelessness: A comparative analysis of common techniques for the simulation of careless responder data* [Unpublished Manuscript]. https://doi.org/https://doi.org/10.31234/osf.io/ge6fa

DeSimone, J. A., DeSimone, A. J., Harms, P. D., & Wood, D. (2018). The Differential Impacts of Two Forms of Insufficient Effort Responding. *Applied Psychology*, *67*(2), 309–338. https://doi.org/10.1111/apps.12117

DeSimone, J. A., & Harms, P. D. (2018). Dirty Data: The Effects of Screening Respondents Who Provide Low-Quality Data in Survey Research. *Journal of Business and Psychology*, *33*(5), 559–577. https://doi.org/10.1007/s10869-017-9514-9

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171–181. https://doi.org/10.1002/job.1962

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In *Multiple Classifier Systems* (Vol. 1857, pp. 1–15). Springer. https://doi.org/10.1007/3-540-45014-9_1

Dupuis, M., Meier, E., Capel, R., & Gendre, F. (2015). Measuring individuals' response quality in self-administered psychological tests: An introduction to Gendre's functional method. *Frontiers in Psychology*, *6 - 2015*, 629. https://doi.org/10.3389/fpsyg.2015.00629

Dupuis, M., Meier, E., & Cuneo, F. (2018). Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices. *Behavior Research Methods*, *51*, 2228–2237. https://doi.org/10.3758/s13428-018-1103-y

Falk, C. F., Huang, A., & Ilagan, M. J. (2025). Unsupervised [randomly responding] survey bot detection: In search of high classification accuracy. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000746

Figurska, M., Stańczyk, M., & Kulesza, K. (2008). Humans cannot consciously generate random numbers sequences: Polemic study. *Medical Hypotheses*, *70*(1), 182–185. https://doi.org/10.1016/j.mehy.2007.06.038

Francavilla, N. M., Meade, A. W., & Young, A. L. (2019). Social Interaction and Internet-Based Surveys: Examining the Effects of Virtual and In-Person Proctors on Careless Response. *Applied Psychology*, *68*(2), 223–249. https://doi.org/10.1111/apps.12159

Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *Leadership Quarterly*, *31*(4). https://doi.org/10.1016/J.LEAQUA.2020.101384

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Gonzalez, O., Canning, J. R., Smyth, H., & MacKinnon, D. P. (2020). A Psychometric Evaluation of the Short Grit Scale. *European Journal of Psychological Assessment*, *36*(4), 646–657. https://doi.org/10.1027/1015-5759/a000535

Goretzko, D., & Bühner, M. (2020). One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis. *Psychological Methods*, *25*(6), 776–786. https://doi.org/10.1037/met0000262

Grau, I., Ebbeler, C., & Banse, R. (2019). Cultural Differences in Careless Responding. *Journal of Cross-Cultural Psychology*, *50*(3), 336–357. https://doi.org/10.1177/0022022119827379

Gummer, T., Roßmann, J., & Silber, H. (2018). Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation. *Sociological Methods & Research*, *50*(1), 238–264. https://doi.org/10.1177/0049124118769083

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2025). Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: Factors influencing careless responding. *Psychological Methods*, *30*(2), 374–392. https://doi.org/10.1037/met0000580

Hong, M., Steedle, J. T., & Cheng, Y. (2020). Methods of Detecting Insufficient Effort Responding: Comparisons and Practical Recommendations. *Educational and Psychological Measurement*, 80(*2)*, 312–345. https://doi.org/10.1177/0013164419865316

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting Insufficient Effort Responding with an Infrequency Scale: Evaluating Validity and Participant Reactions. *Journal of Business and Psychology*, *30*(2). https://doi.org/10.1007/s10869-014-9357-6

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and Deterring Insufficient Effort Responding to Surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

Ilagan, M. J., & Falk, C. F. (2023). Supervised Classes, Unsupervised Mixing Proportions: Detection of Bots in a Likert-Type Questionnaire. *Educational and Psychological Measurement*, *83*(2), 217–239. https://doi.org/10.1177/00131644221104220

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A Practical Guide to Variable Selection in Structural Equation Modeling by Using Regularized Multiple-Indicators, Multiple-Causes Models. *Advances in Methods and Practices in Psychological Science*, *2*(1), 55–76. https://doi.org/10.1177/2515245919826527

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, *10*(JUN). https://doi.org/10.3389/fpsyg.2019.01258

Kam, C. C. S., & Meyer, J. P. (2015). How Careless Responding and Acquiescence Response Bias Can Influence Construct Dimensionality: The Case of Job Satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Karabatsos, G. (2003). Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics. *Applied Measurement in Education*, *16*(4), 277–298. https://doi.org/10.1207/S15324818AME1604_2

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika*, *45*(2), 527–563. https://doi.org/10.1007/s41237-018-0063-y

Leiner, D. J. (2019). Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods*, *13*(3). https://doi.org/https://doi.org/10.18148/srm/2019.v13i3.7403

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December 15). Isolation forest. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 413–422. https://doi.org/10.1109/ICDM.2008.17

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*(1), 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random respondents. *Personality and Individual Differences*, *84*(8), 79–83. https://doi.org/10.1016/j.paid.2014.08.021

Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills*, *112*(2), 331–348. https://doi.org/10.2466/03.11.PMS.112.2.331-348

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Myöhänen, A. (2021). *Fraudulent Respondent Recognition in Adolescents' Health Survey Data Comparison of Data Analysis Methods* [Master's Thesis, Tampere University]. https://trepo.tuni.fi/bitstream/handle/10024/135644/Myo%CC%88ha%CC%88nenAnna.pdf?sequence=2

Nimon, K., Zientek, L. R., & Henson, R. K. (2012). The Assumption of a Reliable Instrument and Other Pitfalls to Avoid When Considering the Reliability of Data. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00102

Olson, K., & Parkhurst, B. (2013). Collecting Paradata for Measurement Error Evaluations. In F. Kreuter (Ed.), *Improving Surveys with Paradata* (pp. 43–72). Wiley. https://doi.org/10.1002/9781118596869.ch3

Open-Source Psychometrics Project. (2020). *Open-Source Psychometrics Project.* Retrieved from https://openpsychometrics.org/

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Ounacer, S., El Bour, H. A., Oubrahim, Y., Ghoumari, M. Y., & Azzouazi, M. (2018). Using Isolation Forest in anomaly detection: the case of credit card transactions. *Periodicals of Engineering and Natural Sciences*, *6*(2), 394–400. htpps://doi.org/ 10.21533/pen.v6i2.533

Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Science and Management*, *4*, 19–29. https://doi.org/10.1016/j.dsm.2021.12.002

R Core Team. (2024). *R: A language and environment for statistical computing* (4.4.0). R Foundation for Statistical Computing.

Schimmack, U. (2020). A meta-psychological perspective on the decade of replication failures in social psychology. *Canadian Psychology*, *61*(4), 364–376. https://doi.org/10.1037/cap0000246

Schroeders, U., Schmidt, C., & Gnambs, T. (2022). Detecting Careless Responding in Survey Data Using Stochastic Gradient Boosting. *Educational and Psychological Measurement*, *82*(1), 29–56. https://doi.org/10.1177/00131644211004708

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. https://doi.org/10.1007/978-0-387-21706-2

Wang, M. D., & Hau, K.-T. (2024). Flagging insufficient effort responses in surveys: Stopping rule to prevent insufficient or excessive removal of doubtful data. *Acta Psychologica*, *243*, 104135. https://doi.org/10.1016/j.actpsy.2024.104135

Ward, M. K., & Meade, A. W. (2023). Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices. *Annual Review of Psychology*, *74*(1). https://doi.org/10.1146/annurev-psych-040422-045007

Woods, C. M. (2006). Careless Responding to Reverse-Worded Items: Implications for Confirmatory Factor Analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3). https://doi.org/10.1007/s10862-005-9004-7

Yentes, R. D. (2020). *In Search of Best Practices for the Identification and Removal of Careless Responders* [Doctoral dissertation, North Carolina State University]. NC State Repository. http://www.lib.ncsu.edu/resolver/1840.20/37220

Table 2.1

Overview of Simulated Scale Characteristics

Scale	Reference	No. items	No. factors	No. Categories	*N*	% Missing
Experiences in Close Relationships	Brennan et al. (1998)	36	2	5	51,491	0.01
Big-Five Inventory	Goldberg (1992)	50	5	5	19,719	0.00
HEXACO Personality Inventory	Ashton et al. (2007)	240	6	7	22,786	0.00

Note. The Likert scale might be different than the one used in the validation study of these scales. Datasets were obtained from *openpsychometrics.org*

Table 2.2

Overview of the Data-Driven IER Indicators

Indicator	Description	Targets	Included in Study 1a	Included in Study 1b, 2 & 3	Cutoff used*	Reference
Long String Index	Longest sequence of identical responses	Excessive invariance	Yes	Yes	x > # items X 0.5	Johnson (2005)
Mahalanobis Distance	Distance of a response pattern from the multivariate center of all response patterns	Multivariate outliers	Yes	Yes	$p < .05$	Meade & Craig (2012)
Person-Total Correlation	Correlation between a respondent' vector of responses and the mean of other respondents' responses	Consistency	Yes	Yes	$x \leq .000$	Karabatsos (2003)
Psychometric Synonyms and Antonyms	Within-person correlation between pairs of highly correlated items	Consistency	Yes	Yes	$x \leq .000$	Meade & Craig (2012)
Intra-individual variability	Standard deviation across a set of responses for a given respondent.	Excessive variance/ invariance	No	Yes	N/A	Marjanovic et al. (2015)
Response Coherence	Predictability of the responses provided by a respondent	Consistency	No	Yes	N/A	Dupuis et al. (2015)
Response Reliability	Stability of the responses provided by a respondent	Consistency	No	Yes	N/A	Dupuis et al. (2015)

Note. *Cutoffs used for the sequential approach only

Table 2.3

Summary of the Performance Measures

Performance Indicator	Formula
Sensitivity/Recall	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Balanced Accuracy	$\frac{Sensitivity + Specificity}{2}$

Note. *TN* = true negatives; *TP* = true positives; *FP* = false positives; *FN* = false negatives; *N* = sample size.

Figure 2.1

Study 1a – Area under the Curve of the IERI with and without a Cutoff

Note. Sample Size: $N = 2000$. A value closer to .50 denotes a weaker discriminative potential. A value closer to .00 or 1.00 denotes a stronger discriminative potential. The cutoffs used for each IERI are specified in *Table 2.*

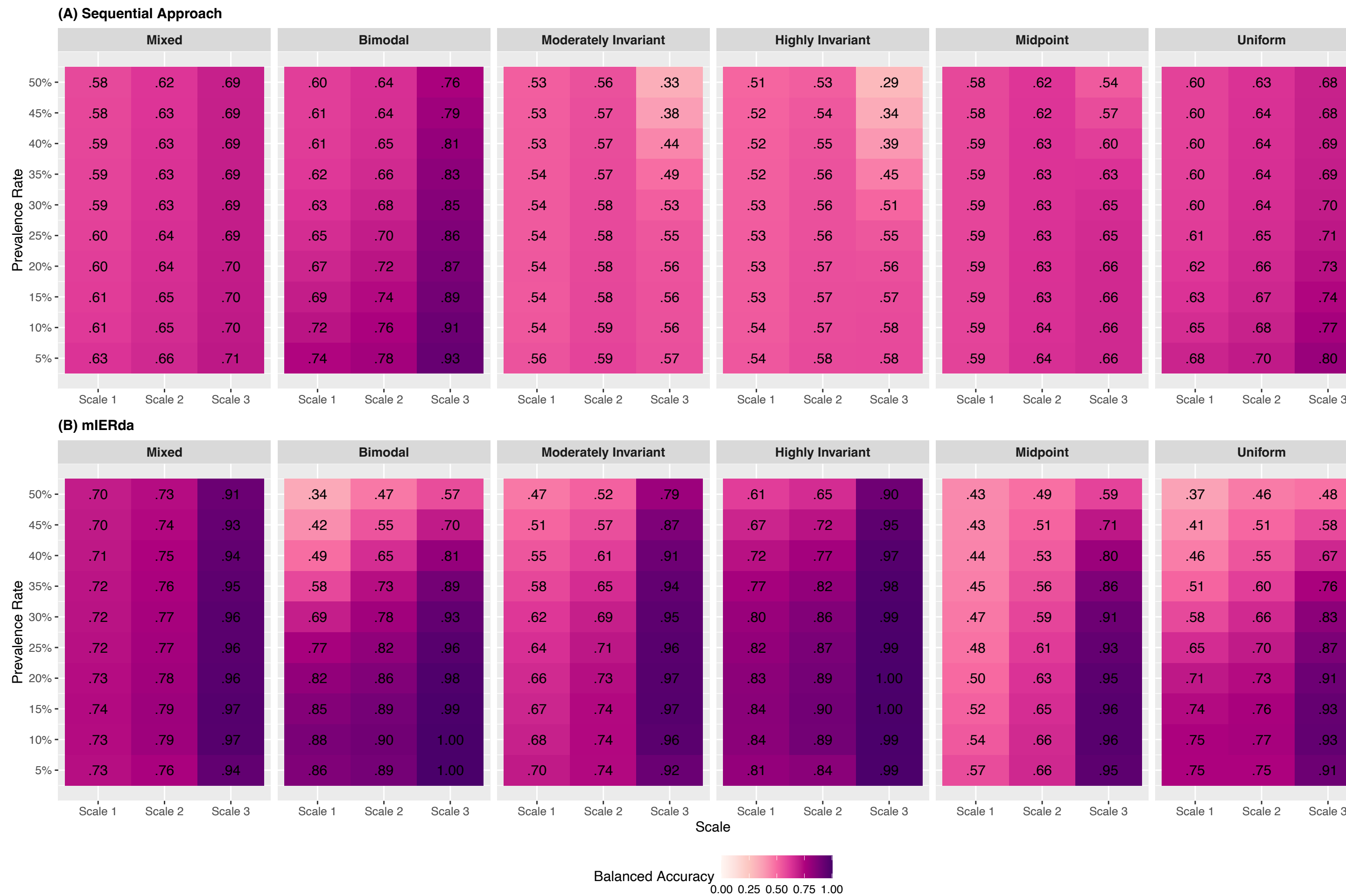
Figure 2.2

Study 1a – Area under the Curve of Additional IERI

Note. Sample Size: $N = 2000$. A value closer to .50 denotes a weaker discriminative potential. A value closer to .00 or 1.00 denotes a stronger discriminative potential. No cutoffs were applied to these IERI.

Figure 2.3

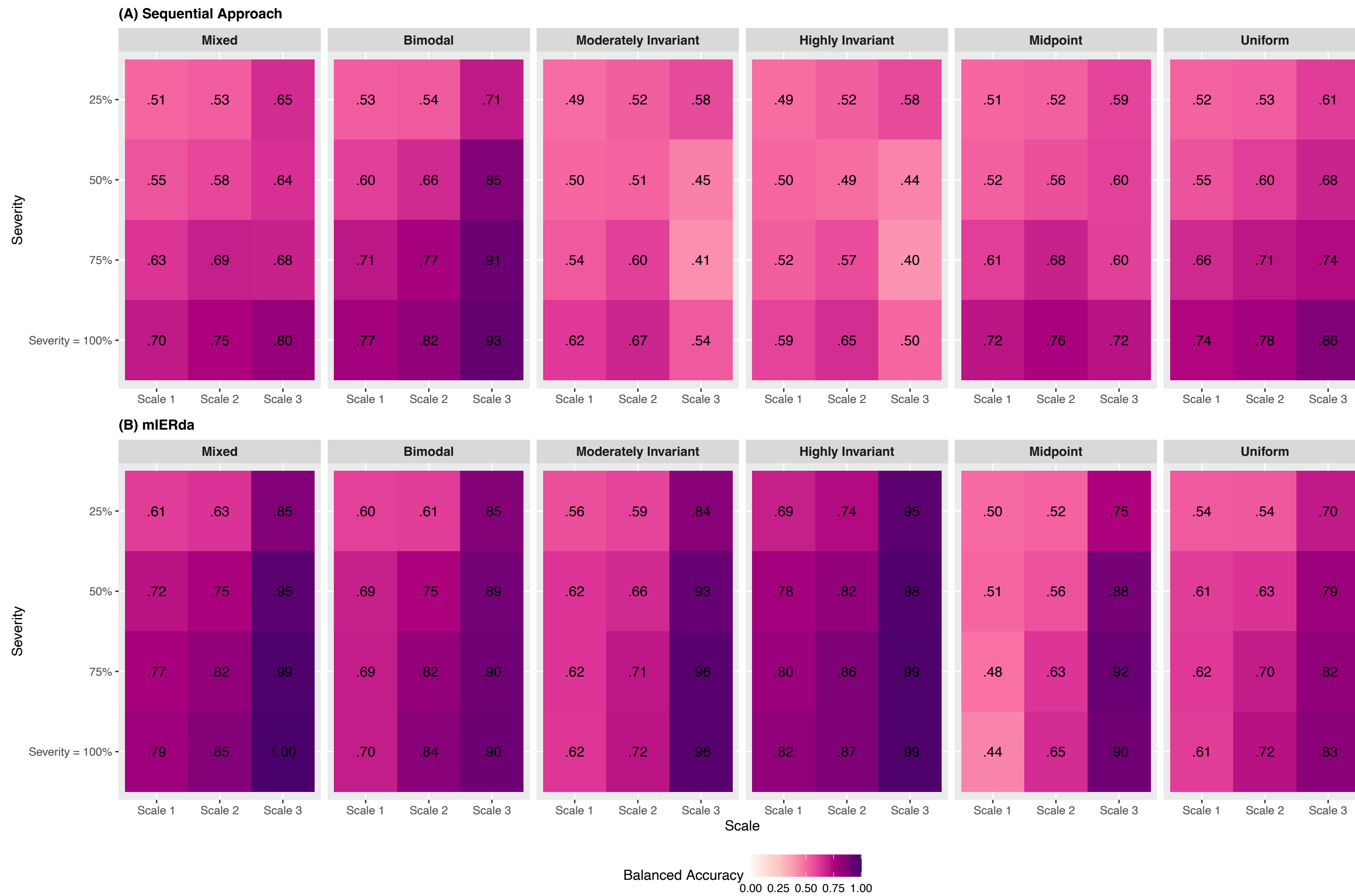
Study 1a – Balanced Accuracy of the Sequential Approach and the mIERda Across Scales and Prevalence Rates



Note. Sample Size: $N = 2000$. A value closer to 1.00 denotes a stronger accuracy.

Figure 2.4

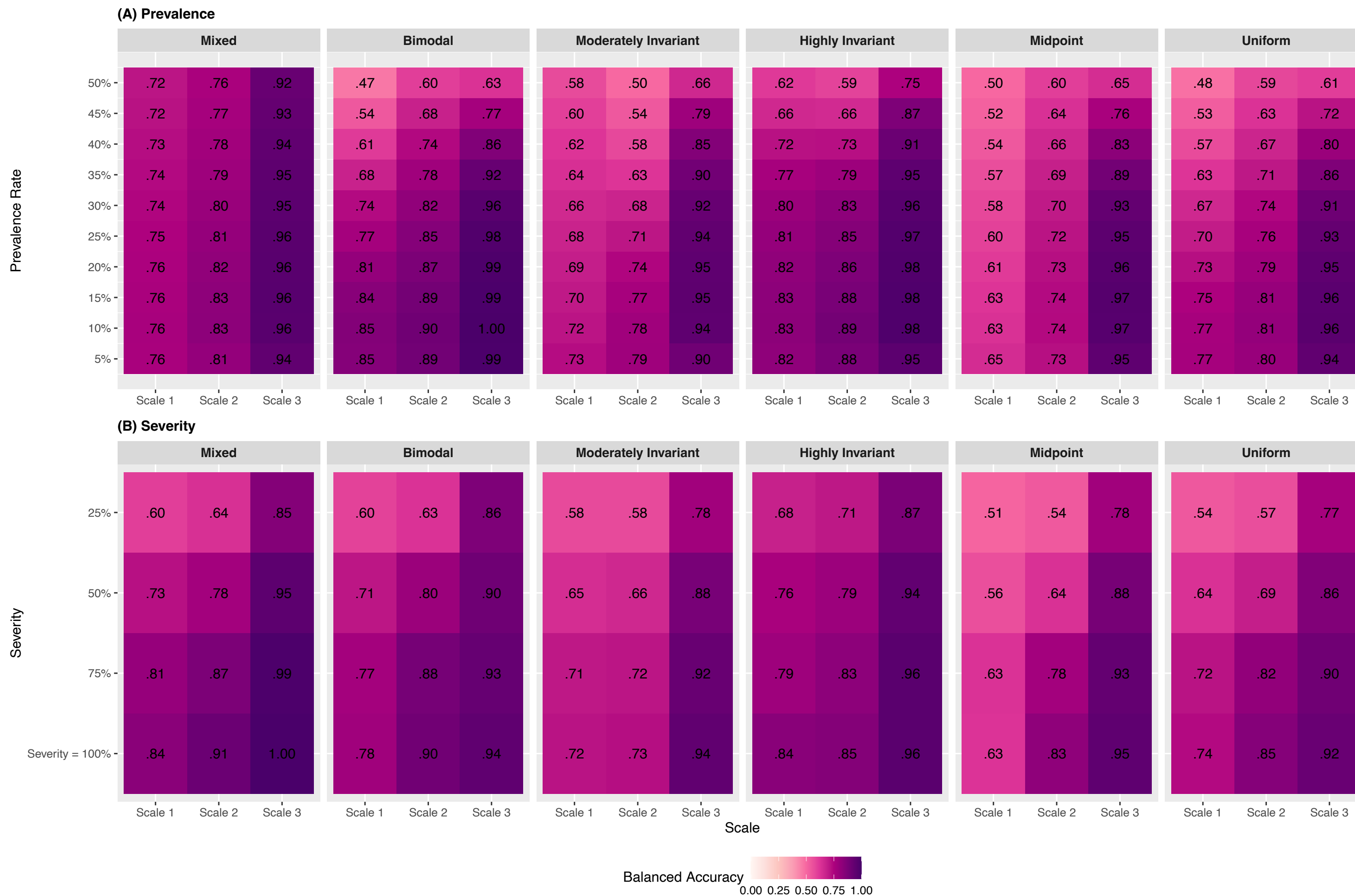
Study 1a – Balanced Accuracy of the Sequential Approach and the mIERda Across Scales and Severity Levels



Note. Sample Size: $N = 2000$

Figure 2.5

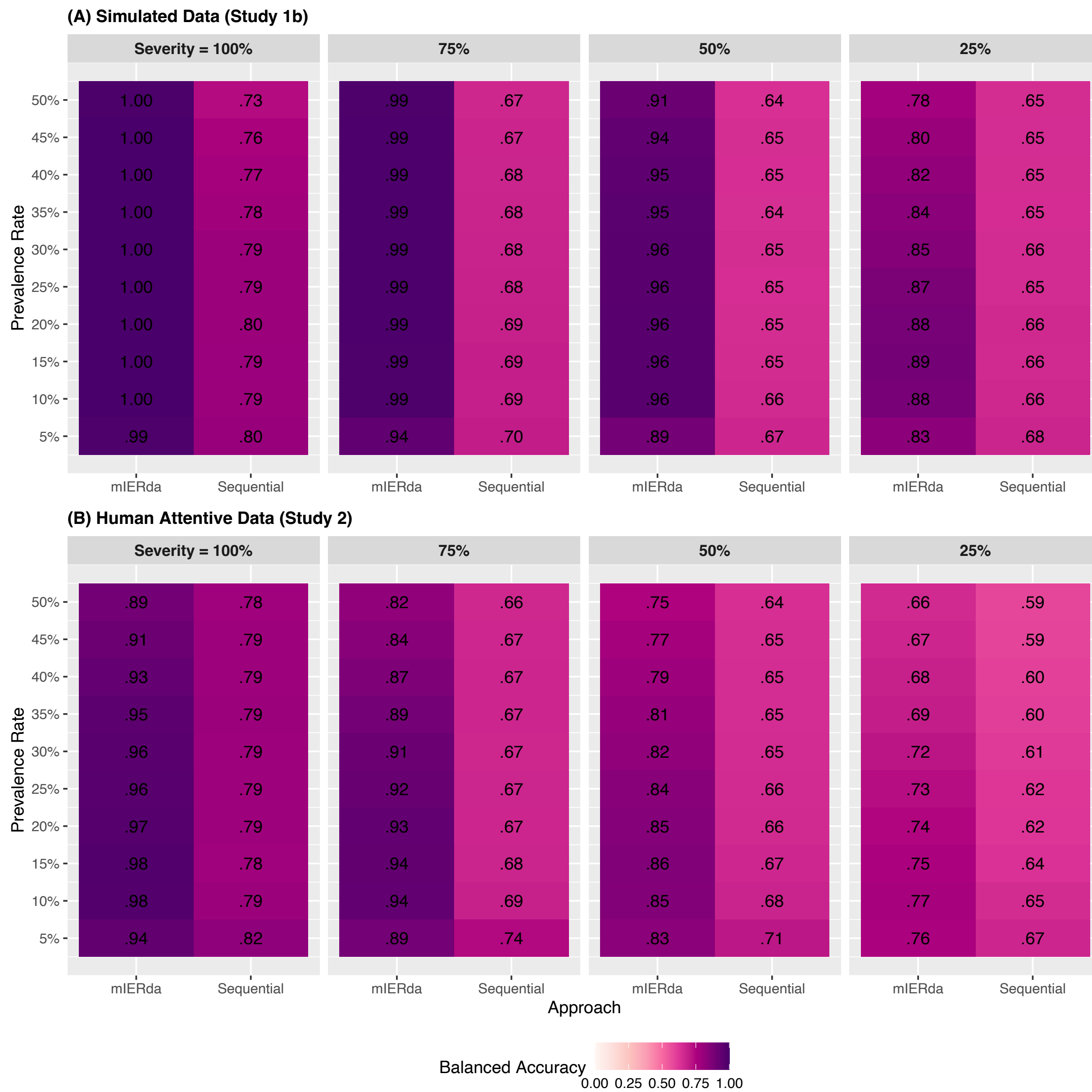
Study 1b – Balanced Accuracy of the mIERda Across Scales, Prevalence and Severity Rates



Note. Sample Size: $N = 2000$. Panel A = Balanced Accuracy by prevalence rate. Panel B = Balanced Accuracy by severity.

Figure 2.6

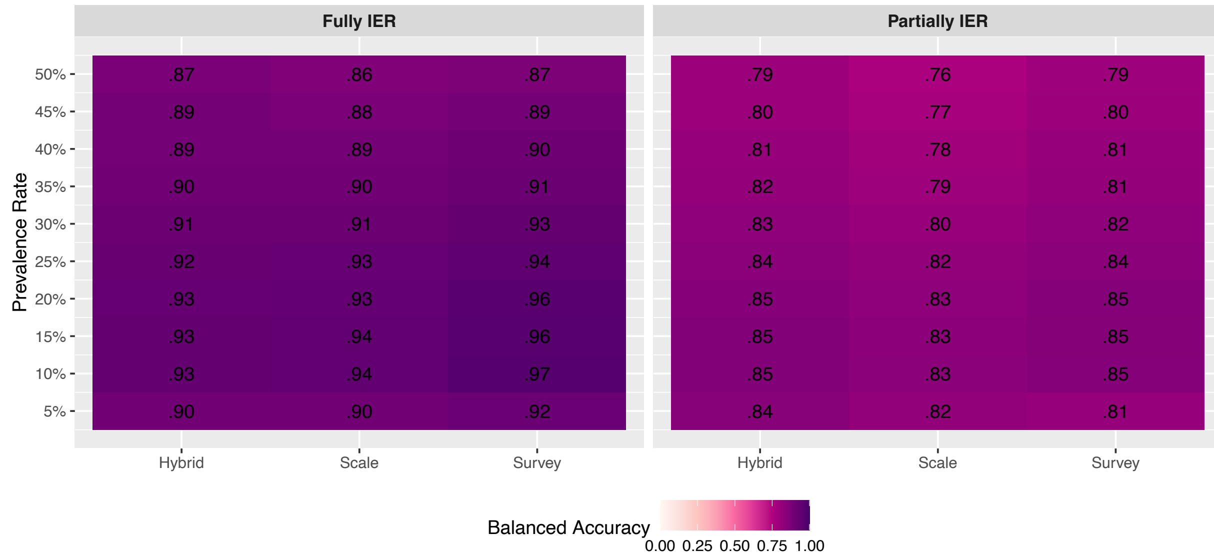
Study 2 – Balanced Accuracy of the mIERda for Mixed Respondents Across Prevalence and Severity Rates for the HEXACO (Scale 3)



Note. Error bars denote the Monte Carlo Standard Errors. To facilitate the interpretation, results from Study 1b were extracted for the same conditions.

Figure 2.7

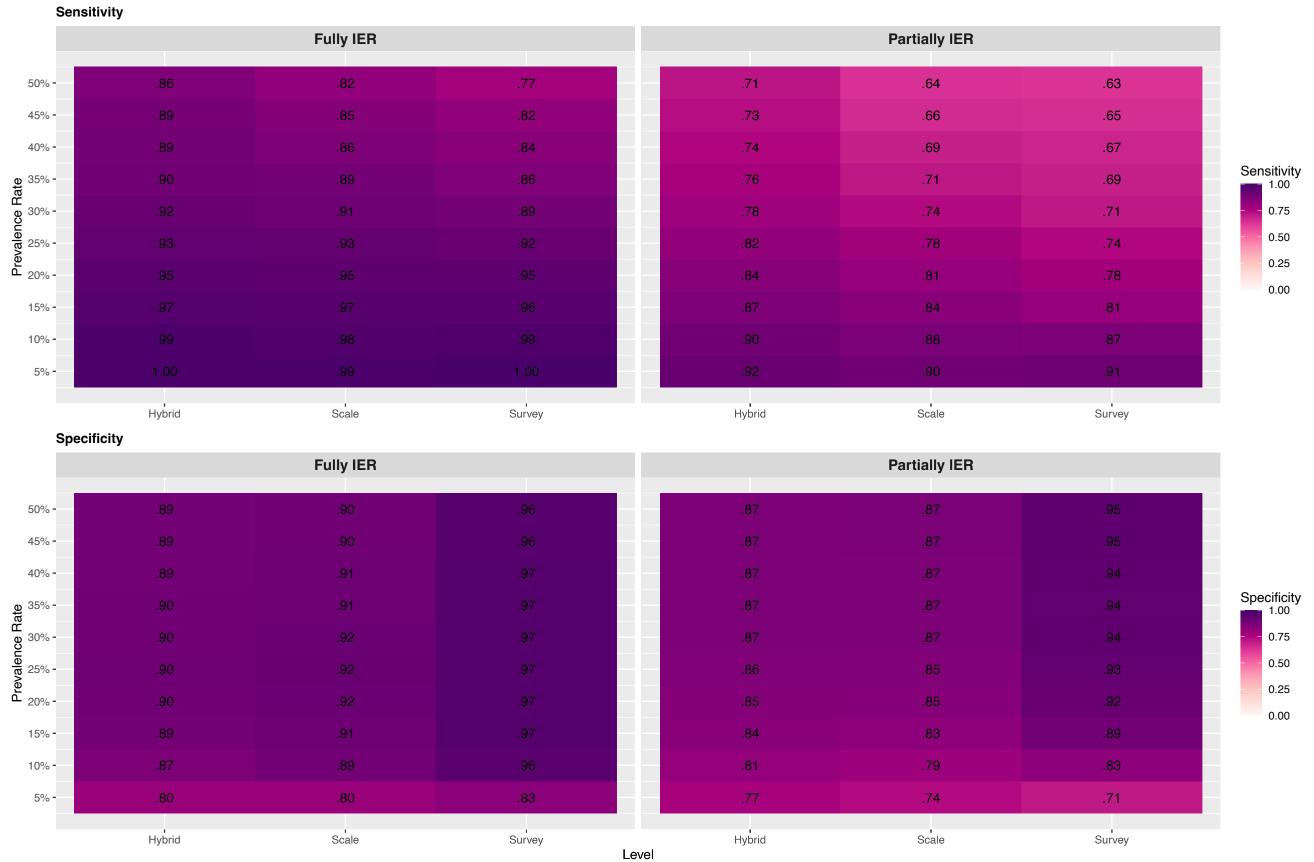
Study 3 – Balanced Accuracy of the mIERda for Mixed Respondents Across Prevalence and Severity Rates for the ECR (Scale 1) and BFI (Scale 2)



Note. N = 500. Pattern = mixed respondents.

Figure 2.8

Study 3 – Sensitivity and Specificity of the mIERda for Mixed Respondents Across Prevalence, Severity Rates and Levels of Screening



Note. $N = 500$

CHAPTER III

Multi-Study Examination of the Construct Validity of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

Ariane J. Gauthier[1] & Luc G. Pelletier[1]

Affiliation
[1]School of Psychology, University of Ottawa
136 Jean-Jacques-Lussier, Ottawa, Ontario, Canada K1N 9A8

Author's Contributions Statement CRediT

Ariane J. Gauthier: Conceptualization (lead), Methodology, Validation, Formal Analysis, Data Curation, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing (lead), Visualization
Luc G. Pelletier: Conceptualization (supporting), Writing – Review & Editing (supporting)

Abstract

Although a large body of research is recognizing the relevance of addressing insufficient effort responding (IER) in survey research, the identification of IER participants remains a challenge. In order to facilitate their identification, Gauthier and colleagues (2025) have introduced a novel approach to IER detection: the multidimensional IER detection approach (mIERda). Using 18 datasets ($N = 5014$) arising from five distinct research programs, the present research examines the nomological network of the mIERda and provides evidence of its convergent, discriminant, predictive, concurrent, and known-groups validity throughout the survey. Results from pooled correlations analyses provided evidence of the convergent validity of the mIERda as demonstrated by its significant correlation with other indicators of IER. The mIERda also showed significant correlations with known contextual (i.e., time of the semester) and personal determinants (i.e., amotivation and autonomous motivation at a global and research level, conscientiousness) of IER associated with IER in past research, supporting its concurrent and predictive validity. Results also supported the discriminant and known-groups validity of the mIERda as showed, respectively, by a null correlation between the mIERda and unrelated correlates of IER (i.e., controlled motivations and social desirability) and a greater prevalence of IER in men than in women. Taken together, this research provides evidence of the construct validity of the mIERda throughout the survey (i.e., at the start and the end of the survey), positioning it as a valid and encompassing IER detection tool.

Keywords: insufficient effort responding, careless responding, data quality, validation, survey research.

Multi-Study Examination of the Construct Validity of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

As researchers conduct research using surveys, ensuring the quality of said data is of upmost importance. When researchers collect data using online surveys, they rely on the assumption that their participants respond honestly, truthfully and at the very least, attentively. Results arising from a growing body of literature suggest that 3% to 50% of participants are disengaged when they partake in survey research (Francavilla et al., 2019; Meade & Craig, 2012), a phenomenon better known as insufficient effort responding (IER). IER is defined as a lack of motivation to comply with survey instructions, interpret, and provide accurate responses to items (Huang et al., 2012). What stems from participants' disengagement can impact research results to the point of distorting factorial structure (Arias et al., 2020; Huang, Liu, & Bowling, 2015; Kam, 2019; Woods, 2006), and diminishing statistical power, construct validity and effect sizes (Kam & Meyer, 2015; Maniaci & Rogge, 2014; Oppenheimer et al., 2009). IER, consequently, represents a considerable threat to the validity of conclusions drawn from survey research.

Even if the potential consequences of IER are well-documented, and the scientific community is increasingly recognizing the importance of addressing IER in data drawn from survey research, the identification of these participants is challenging. Without the right indicators, IER often goes undetected (Curran, 2016). As such, multiple indicators of IER (e.g., attention checks, self-reported data quality items, completion time) have been introduced to facilitate the identification of IER. (Curran, 2016; Dupuis et al., 2018; Marjanovic et al., 2015; Meade & Craig, 2012; Schroeders et al., 2022) and researchers have been encouraged to rely on a screening process known as the sequential approach (Curran, 2016; Ward & Meade, 2023). To

implement the sequential approach, researchers must select multiple indicators of IER and establish cutoffs for IER screening. Nonetheless, most of past research pertaining to the identification of IER has focused on the development of indicators rather than the implementation of said cutoffs (Dupuis et al., 2018; Marjanovic et al., 2015; Meade & Craig, 2012). Consequently, little is known about which indicator should be used and the optimal cutoffs for these indicators, which hinders the quality of the screening process. To address the complexities inherently tied to the use of the sequential approach, there is a need to move beyond cutoffs and to introduce an easily implementable data-driven approach.

A Novel Approach to IER Detection: the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

The mIERda is a novel IER detection approach which seeks to improve the accuracy and the quality of the IER screening performed by researchers. Developed by Gauthier et al. (2025), the mIERda relies on various data-driven indicators of IER (i.e., intra-individual variability, long string index, Mahalanobis Distance, person-total correlation, response coherence, and response reliability). Whereas the typical application of these indicators requires the researcher to establish a threshold beyond which one should be considered IER for each indicator, the mIERda offers a novel perspective. Rather than forcing researchers to impose cutoffs, then to sequentially and individually interpret these indicators, the mIERda defers their interpretation to an adaptation of an unsupervised learning algorithm (i.e., fair-cut forest). By doing so, the raw scores of the indicators are directly treated by the algorithm underlying the mIERda which not only relieves researchers from the need to establish arbitrary cutoffs but also maximizes the information provided by each indicator (Gauthier et al., 2025).

Results from a series of simulation studies found evidence of the criterion validity of the mIERda by examining its classification accuracy under various conditions (Gauthier et al., 2025). Results from a first study showed that the IER screening performed by the mIERda correctly classified more participants as either attentive or IER than the sequential approach, which is the current gold standard in terms of IER detection. In a second study, it was found that the mIERda maintained a high level of accuracy when screening simulated IER responses that were added amongst human-generated data. The mIERda was once again found to be more accurate than the sequential approach under such circumstances. A third study found that the classification accuracy of the mIERda improved when screening multiple scales simultaneously, supporting its application in a multiple-scale screening setting. Gauthier and colleagues (2025) demonstrated that the mIERda was more accurate when used to screen IER among longer scales or on more items, especially when the data contained a mixture of random and non-random IER participants. Moreover, when simulated participants engaged in IER on all survey items, the mIERda revealed superior accuracy relative to instances where IER was limited to a specific item subset.

While these results provide initial evidence of the criterion validity of the mIERda, its construct validity remains to be established. Achieving this goal requires us to examine whether the mIERda behaves as expected within the nomological network of IER.

Nomological Network of Insufficient Effort Responding

Given the pervasive impact of IER on survey research results, researchers have been interested in understanding the factors and correlates of IER in the hope of preventing its occurrence. The literature frames IER as resulting from various contextual and personal factors that seem to foster or deter participants from engaging in IER.

Contextual Factors

Contextual factors refer to all the circumstantial elements that define the context in which the survey takes place. Past research has investigated the role of various contextual determinants of IER. Notably, the time of the semester and the length of the survey are two widely accepted contextual determinants of IER. More precisely, past research has found that participants who complete surveys later in the semester tend to engage more often in IER than participants who complete surveys earlier in the semester (Aviv et al., 2002; Cheung, 2019; Witt et al., 2011). The exact cause of this finding is unknown. However, it is posited that participants might be more fatigued towards the end of the semester, which could make them more prone to IER (de Calvo & Reich, 2007). Similarly, results from past research suggest that IER might be impacted by the length of the survey in a way where longer surveys are associated with a greater prevalence of IER (Bowling et al., 2021; Galesic & Bosnjak, 2009; Gibson & Bowling, 2020; Ward et al., 2017). Results from a study posit that although some participants engage in IER at the start of a survey, IER becomes more prevalent towards its end, especially in longer surveys (Berry et al., 1992). Besides the literature suggesting that IER is more common in longer surveys, Bowling and colleagues (2021) revealed that IER increases at a within-person as they progress through the survey. Such results, combined to Berry and colleagues' (1992) findings suggesting that IER is more common at the end of the survey especially in longer surveys, support the relevance of not only examining survey length as a correlate of IER but, also, of examining whether the correlates of IER differ as participants progress within a survey.

Personal Factors

Besides contextual factors, IER is also associated with various personal factors. Gender, personality, and motivation are three personal factors which tend to have a relatively stable

association with IER. More precisely, past research has shown that men are more prone to engage in IER than women (Braitman et al., 2025; Cheung, 2019; Schneider et al., 2018). Besides gender, the association between personality and IER has been repeatedly reported in the literature. Results from past research suggest that higher levels of extraversion, conscientiousness, agreeableness, and emotional stability were associated with lower incidence of IER (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017), and that IER tend to be more strongly and consistently associated with conscientiousness (Mazza & Arthur, 2025). In addition to personality, researchers have also studied the association between IER and motivation. Such investigation is warranted given that IER is thought to be caused by a lack of motivation to read, understand, or respond to survey items (Huang et al., 2012). Past research provided evidence that motivation, or lack thereof, is associated with a higher prevalence of IER (Cheung, 2019; Hasselhorn et al., 2025). More precisely, Cheung (2019) has investigated the association between IER and various types of motivation, as defined by Self-Determination Theory (SDT; Deci & Ryan, 2017), in samples of participants recruited from a university-based research subject recruitment pool. Through their work, they have demonstrated that global (i.e., overall disposition), academic (i.e., toward academic tasks), and research (i.e., toward research participation) amotivation share a positive association with IER (i.e., completion time, semantic synonyms and antonyms, long string index; Cheung, 2019). This means that participants with greater levels of amotivation tended to engage more often in IER. As well, Cheung (2019) demonstrated that participants with greater level of autonomous motivation (i.e., at a global, academic, and research level) displayed lower engagement in IER. Lastly, Cheung (2019) has failed to find any evidence that controlled motivation at a global, academic, and research level

was associated with IER. This suggests that controlled motivation does not seem to have an impact on participants' engagement in IER.

Taken together, these results suggest that different types of motivation seem to have a differential impact on IER. Building on SDT (for an in-depth review of the theory refer to Ryan & Deci, 2017; Ryan, 2023) can allow us to understand these nuances. Within SDT, it is posited that the reasons why individuals engage in various behaviours can be defined along a continuum of autonomy (Ryan & Deci, 2017). Overall, it is stated that acting in a way that is congruent with oneself (i.e. *autonomous motivation*), rather than behaving in a way that feels incongruent with the self (i.e., out of internal or external pressure; *controlled motivation*) or feels meaningless (i.e., *amotivation*), is not only associated with greater well-being, but also greater behavioural persistence. Taken together, as we consider Cheung's (2019) results, this suggests that participants who tend to act in a way that is congruent with the self were less likely of engaging in IER. In opposition, individual who have a stronger disposition to engage in behaviours that feel meaningless to them, or who reported that engaging in research and academic tasks without seeing the purpose of doing so, were more likely of engaging in IER. Interestingly Cheung's (2019) results suggest that engaging in research to address internal or external pressures (e.g. to obtain a compensation) did not impact participants engagement in IER. Overall, these differential associations highlight the relevance of considering different types of motivation, or lack thereof, when examining the nomological network of IER.

Limitations of Existing Research

While Gauthier and colleagues (2025) have introduced the mIERda and examined its criterion validity through a series of simulation studies, further research is needed to expand the validation of the mIERda. Currently, it remains unknown how the mIERda behaves within its

broader nomological network. More specifically, how the mIERda relates to other measures of IER (i.e., convergent validity), measures of unrelated constructs (i.e., discriminant validity), antecedents of IER measured prospectively (i.e., predictive validity) or concurrently (i.e., concurrent validity), and across groups known to differ in terms of prevalence of IER (i.e., known-groups validity) remains to be ascertained. Considering the mIERda as a new method for measuring IER, assessing its construct validity is an essential step to ascertain that it measures the construct it is intended to.

Current Study

The current research sought to establish the construct validity of the mIERda across 18 distinct datasets arising from five distinct research programs from the discipline of psychology (for a comprehensive overview of the objective of each research program and the content of each dataset, refer to Appendix C). To achieve this goal, a series of pooled correlation analyses were conducted. More precisely, the convergent validity of the mIERda was assessed using a series of correlation analyses to assess whether the IER status, as defined by the mIERda, relates with other measures of IER which are not included in the mIERda (i.e., attention checks, self-reported data quality items, and completion time). The discriminant validity was assessed by examining the correlation between the mIERda and variables that should not be related to IER. More precisely, we examined the association between the mIERda and social desirability. Then, building on past research suggesting a null association between IER and controlled motivation (Cheung, 2019), we assessed the correlation between the mIERda and controlled motivation (at a global and research level). The predictive validity of the mIERda was assessed by analyzing correlations between the mIERda and established IER correlates, which were assessed independently from IER. To this end, the association of the mIERda with amotivation (at a

global and research level), autonomous motivation (at a global and research level), and conscientiousness was examined. As it pertains to the concurrent validity, it was ascertained by estimating the correlation of the mIERda with the time of the semester at which the participant completed the survey, and the survey length.

Hypotheses

Based on past research, we hypothesized the nomological network presented in Figure 3.1. More precisely, based on Meade and Craig (2012), it was hypothesized that results would provide evidence of the convergent validity of the mIERda as demonstrated by a positive correlation with other measures of IER. We expected at least a moderate correlation ($r \leq .30$) of the mIERda with (1a) attention checks. We also expected a weak to moderate ($r \leq .10$ - $r \geq .50$) correlation of the mIERda with (1b) self-reported data quality items, as well as a small correlation ($r \leq .10$) with (1c) completion time. As it regards discriminant validity, it was hypothesized that the mIERda would be statistically unrelated to (2a) social desirability (Meade & Craig, 2012), controlled (2b) global and (2c) research motivation (Cheung, 2019). Regarding predictive validity, it was hypothesized based on results from Cheung (2019) that the mIERda would have a weak and positive correlation ($r \leq .10$) with (3a) global and (3b) research amotivation. Conversely, we expected the mIERda to have a small negative correlation ($r \leq -.10$) with autonomous (3c) global and (3d) research motivation (Cheung, 2019), as well as with (3e) conscientiousness (Arias et al., 2020; Bowling et al., 2016). As it pertains to the concurrent validity of the mIERda, it was hypothesized that (4a) survey length (Galesic & Bosnjak, 2009; Gibson & Bowling, 2020; Ward et al., 2017) and (4b) time of the semester (Cheung, 2019; Witt et al., 2011) would share a weak positive correlation with IER throughout the survey. Lastly, it was hypothesized that (5) the prevalence of IER would be greater in men than in women

(Braitman et al., 2025; Cheung, 2019; Schneider et al., 2018; Witt et al., 2011), which would support the known-groups validity of the mIERda.

Method

Participants

A total of 5014 participants were recruited across the 18 studies included in the analyses. Respondents were recruited through an eastern Canadian university research subject recruitment pool. Respondents were students enrolled in introductory or second-year courses with a research participation component in psychology, business management, communications, or linguistics. Depending on the length of the survey, respondents received 0.5% to 1% towards their final grade as a compensation for their participation. At the moment of participation, the median age at participation was 19 years old (n = 1253, range = 16 – 31+). Most respondents identified as women (63.7%; n = 3196), while 25.5% identified as men (n = 1279), 6.6% identified otherwise (n = 332), and 4.1% did not disclose their gender (n = 207). Most respondents identified as Caucasian (n = 2104; 42.0%), Asian (n = 1018; 20.3%), or Black (n = 819; 16.3%). For a sample-wide overview of the sociodemographic statistics, consult Table 3.1. Otherwise, refer to Appendix C for a summary of this information by research program. Data from this research arise from a larger research project. The participants included in this study are the ones who appeared once in the dataset (i.e., solely completed one study).

Procedures and Measures

A total of 16 validation variables were used to assess the construct validity of the mIERda. Given the number of considered variables, the measures are summarized in Table 3.2 and presented in more details in Appendix A.

To limit respondents' burden and selection bias, the present research relies on secondary use of data. As they signed up for the subject pool, participants responded to a pre-screen questionnaire which aimed to collect sociodemographic information (e.g., gender) and contained questions pertaining to respondents' motivation for research participation, global motivation, and personality. The complete version of these scales can be found in Appendix B. Upon completion of the pre-screen questionnaire, participants gained access to the subject pool and were free to sign-up and participate in any study of their choosing during the semester. The present research analyzes datasets from studies conducted between the Fall 2021 and Winter 2024 semesters. The datasets used in this research do not originate from studies with exclusion criteria. Before signing up, participants were provided with a summary of the study, its estimated duration and the compensation involved. After signing up for any studies included in the present research, respondents were given access to an online survey, hosted on Qualtrics, to be completed at a moment of their choosing. Their responses to the survey were used to measure IER. It is also worth mentioning that the surveys contained attention checks (i.e., "*I eat cement occasionally*") and self-reported measures of data quality (i.e., self-reported effort, attention, and data inclusion). The datasets included in this research were collected as part of five distinct research programs: academic resilience, the use of cellphone in the classroom, motivation and well-being, physical activity, and pro-environmental behaviours. Datasets were found to have the same topic is they arose from the same research program and research laboratory, thus pertained to the same main topic. A summary of the objectives, the number of datasets and characteristics from each research program appears in Appendix C (Table C-3.1). The specific scales included in each research program, and dataset, can also be found in Appendix D. The raw version of the datasets

tied to each research program was provided upon agreement, after data collection, by their respective principal investigator.

Analytical Plan

Our analytical strategy followed various steps to explore the validity of IER across numerous datasets.

Step 1 – IER screening using the mIERda. To begin, we screened raw datasets for IER using the methodology outlined by Gauthier and colleagues (2025). This process was conducted in *R* (version 4.4; R Core Team, 2024) using the *mIERda* package (Gauthier, 2025). For each dataset, we performed IER screening using the data from the scales containing the first 50 and last 50 items of the survey. When performing IER screening, we excluded the socio-demographic, True/False, and open-ended items because the mIERda was not designed nor validated for such applications. IER screening was performed at a dataset level. Since some indicators of IER (i.e., response coherence and response reliability) are dependent on scales' factorial structure, we avoided splitting any scales. As such, we included as many scales as needed to get as close to 50 items (range: 47 – 54) without dividing any scales. This was done to obtain a measure of IER, at the start and end of the survey, that was as consistent and comparable as possible. As demonstrated by Gauthier et al. (2025), the number of items used for screening can impact the accuracy of the screening performed. Since the present study relies on datasets of varying lengths, the main objective was to reduce bias and ensure comparable IER screening across datasets by screening a similar number of items. Additionally, such approach also offers the opportunity to validate the mIERda at different points of the survey which is especially relevant given that engagement in IER is a dynamic process, meaning that participants can fluctuate in terms of engagement throughout the survey (Bowling et al., 2021). Such approach

might allow us to assess the validity of the mIERda more thoroughly. Performing IER screening allowed us to retrieve an IER status (0 = attentive, 1 = IER), at the start and the end of the survey, for each participant.

Step 2 – Data Preparation. Once IER screening was completed, metadata was extracted from each dataset. More specifically, the length of the study (i.e., number of survey items), number of weeks since the beginning of the semester, and the study topic were retrieved. Afterwards, respondents' socio-demographic information (i.e., age, gender, ethnicity), motivation towards research, global motivation, and personality were extracted from the pre-screen questionnaire. This data was subsequently linked to each dataset.

Step 3 – Examination of the construct validity of the mIERda. Before examining the validity of the mIERda, we extracted the relevant descriptive statistics. More precisely, the prevalence of IER throughout the survey was examined using frequency analyses and McNemar's chi-squares were performed to examine if the aggregated prevalence of IER differs at the start and end of the survey. Differences in prevalence were also examined across datasets and programs of research. We also examined the sensitivity of the mIERda by estimating the proportion of participants flagged as IER by attention checks and the self-reported data inclusion item that were also flagged by the mIERda.

Afterward, the nomological network of the mIERda was examined by estimating the pooled correlation between the mIERda and external criteria (i.e., other indicators of IER, global and research motivation, conscientiousness, social desirability, time of the semester, survey length and gender). Point-biserial correlations were used to estimate the relationship between numerical validation variables (i.e., self-reported effort and attention, global and research motivation, time of the semester, survey length, social desirability) and the mIERda. The

relationship between dichotomous validation variables (i.e., attention checks, self-reported data inclusion, gender) and the mIERda was estimated using tetrachoric correlations. These correlations were firstly computed at a dataset-level (n = 18). These correlations were then pooled using Fisher's Z transformation to get a more reliable and robust overview of the association of the mIERda (i.e., at the start and the end of the survey) with the validation variables. These correlations were estimated at the start and at the end of the survey. Correlations were not interpreted at the dataset level as the objective was to get a robust and generalizable estimate of the validity of the mIERda. For transparency, the correlations at a dataset level are reported in Appendix C. Given the number of comparisons and datasets included in the present study, the significance of results was interpreted at an alpha-level of 1% ($p < .01$).

Results

Preliminary Analyses

Prior to performing the IER screening, each dataset was formatted and screened for out-of-range values. Each dataset was screened for excessive missingness, and respondents with over 30% missing data were excluded from the analysis. After ensuring the data to be screened was in a numerical format, IER screening was performed as per Gauthier and colleagues' (2025) methodology using the mIERda package (Gauthier, 2025; version 0.4).

Afterwards, we generated an analytical dataset by extracting the relevant metadata from each dataset (i.e., completion time, survey length and time of the semester) and variables from the pre-screen questionnaire (i.e., personality, global motivation, gender). We then conducted missing data screening on the motivation and personality variables extracted from the pre-screen data. Little's MCAR test ($\chi^2(3854) = 22\,641.56$, $p < .001$) revealed that the data was not completely missing at random. Data was thus assumed to be missing at random (MAR) and was

replaced with multiple imputation using predictive mean matching using the MICE package (Buuren & Groothuis-Oudshoorn, 2011).

Descriptive Statistics

The descriptive statistics of all the variables and measures used in the present study can be found in Table 3.3 and Table 3.4. The pooled correlations across datasets between the mIERda and the validation variables feature in Table 3.5. These correlations for each dataset are presented in Appendix C. The sensitivity of the mIERda to participants flagged by attention checks and the self-reported data inclusion item is displayed in Table 3.6.

Using the mIERda, we found an average IER prevalence, across datasets, of 32.5% (range: 14.8% - 50.0%) at the start of the survey and of 35.7% (range: 26.3% - 64.2%) at the end of the survey (Table 3.7). Data revealed that 49.9% of participants engaged in IER at any point throughout the survey. The prevalence of IER was significantly greater at the end of the survey (McNemar's $\chi^2(1, N = 5014) = 13.19$, $p < .001$) than at the start of the survey. The prevalence of IER was also found to differ across datasets at the start ($\chi^2(17, N = 5014) = 55.86$, $p < .001$) and the end of the survey ($\chi^2(17, N = 5014) = 77.67$, $p < .001$). This suggests that the prevalence of IER varied across studies.

Convergent Validity

As displayed in Table 3.5, the pooled correlation analysis revealed that the mIERda shared a moderate-to-strong positive correlation with attention checks and the self-reported data inclusion item at the start and the end of the survey. Such correlations suggest that the mIERda and these indicators of IER measure something similar without reaching redundancy. Furthermore, as demonstrated in Table 3.6, the mIERda flagged as IER 78.3% of participants who failed at least two attention checks, 92.2% of participants who failed at least three attention

checks, and 100% of participants who failed four or more attention checks. Additionally, the mIERda flagged as IER 81.2% of participants who reported that their data should be excluded from the analysis. Such results are particularly interesting if we consider that attention checks solely identified 44.5% of the participants flagged as IER by the mIERda, and that the self-reported data inclusion item identified 17.0% of these IER participants[2]. This suggests that although the mIERda identifies most of the participants flagged by these indicators, these indicators do not identify most participants identified by the mIERda.

The mIERda was also found to have a weaker positive correlation with self-reported effort and self-reported attention items across the survey. Finally, the mIERda was found to have a small negative correlation with survey completion time irrespectively of when the data was screened (i.e. throughout the survey). These findings support hypotheses 1a, 1b, & 1c and provide evidence of the convergent validity of the mIERda at the start and the end of the survey.

Discriminant Validity

Based on past research, it was hypothesized that social desirability, controlled motivation at a global and research level would be unrelated to the mIERda. Results from the pooled correlation analyses indeed demonstrated that these variables were not significantly correlated with the mIERda at the start and the end of the survey. These results support hypotheses 2a, 2b & 2c and provide evidence of the discriminant of the mIERda at the start and the end of the survey.

Predictive Validity

As expected, autonomous motivation at a global- and research-level, and conscientiousness were found to have a small and negative correlation with the mIERda across

---

[2] Amongst participants who self-reported their data should be excluded from analysis, 67.6% failed three or more attentions checks. Amongst participants who failed at least three attention checks, 40.3% self-reported that their data should be excluded from analysis.

the survey. The mIERda was also found to have a weak positive correlation with amotivation at a global and research level irrespectively of the time of the survey. Taken together, these results support our hypotheses (3a, 3b & 3c) and provide evidence of the predictive validity of the mIERda.

Concurrent Validity

As hypothesized, results revealed a weak positive correlation between the mIERda and the time of the semester. This effect was observed at the start and the end of the survey and indicates that the prevalence of IER increased throughout the semester. Such results supported hypothesis 4b. The correlation between the mIERda and the survey length was, however, non-significant at the start of the survey. Surprisingly, the survey length was found to be weakly and negatively correlated with the mIERda at the end of the survey. This suggests that the prevalence of IER, at the end of the survey, was lower in longer surveys. Such effect was contrary to hypothesis 4a.

Given these unexpected results, to assess whether or not this effect is specific to the mIERda, we assessed the correlation between other indicators of IER (i.e., the IER indicators used to assess the convergent validity) and survey length. Results unveiled a negative association of survey length with attention checks ($r_{pooled\ (N = 3865)} = -.03$, $CI_{95}[-.04, -.02]$, $p < .001$) and a null association of survey length with self-reported effort ($r_{(N = 3865)} = -.02$, $CI_{95}[-.05, .01]$, $p = .286$), self-reported attention ($r_{(N = 4178)} = .02$, $CI_{95}[-.01, .05]$, $p = .122$), self-reported data inclusion ($r_{(N = 4203)} = .00$, $CI_{95}[-.03, .03]$, $p = .976$), and survey completion time ($r_{(N = 5014)} = .00$, $CI_{95}[-.03, .03]$, $p = 1.00$). Overall, these results suggest that the null association (at the start of the survey) and negative association (at the end of the survey) observed between survey length and the mIERda persists when using other measures of IER. This implies that this effect was not limited

to the mIERda. Overall, this suggests that although hypothesis 4a was not supported, results offer evidence of the mIERda's concurrent validity.

Known-Groups Validity

As hypothesized, results revealed a positive correlation between the mIERda and participants' gender (0 = Women, 1 = Men) suggesting that men were more likely than women to engage in IER throughout the survey. This supports our hypothesis (5) of the known-groups validity of the mIERda.

Discussion

The mIERda was conceptualized as a novel and encompassing measure of IER which addresses the limitations tied to prior IER detection approaches. Instead of requiring researchers to implement cutoffs and interpret indicators of IER, the mIERda relies on a machine learning algorithm to facilitate the implementation of IER screening. The present study showcases the results of a multi-study examination of the convergent, discriminant, predictive, concurrent and known-groups validity of this novel IER detection approach across 18 datasets. More specifically, we examined the association of the mIERda throughout the survey with other indicators of IER (i.e., attention checks, self-reported effort, attention and data inclusion), known personal predictors of IER (i.e., autonomous motivation and amotivation at a global and research level, conscientiousness), contextual determinants of IER (i.e., time of the semester, survey length), constructs unrelated to IER (i.e., controlled motivation at a global and research level, social desirability). The extent to which the mIERda identifies participants flagged by attention checks and self-reported data inclusion was also assessed.

Validity of the mIERda

Given the dynamic nature of IER throughout the survey, the validity of the mIERda was examined at the start and the end of the survey. Below, we discuss the different forms of validity and their association with the mIERda throughout the survey.

Convergent Validity

The indicators of IER included in the present research were found to correlate, as hypothesized, with the mIERda. Overall, the results support the convergent validity of the mIERda. The strength of these correlations, however, differed across indicators. Results indeed suggest that some indicators of IER were more strongly related to the mIERda than others. For instance, the mIERda was found to share a moderate-to-strong positive correlation with attention checks and the self-reported data inclusion item throughout the survey (i.e., at the start and the end of the survey). Such correlation suggests the mIERda shares significant and meaningful variance with these indicators and that they measure a similar construct. In contrast, the association of the mIERda with remaining indicators of IER (i.e., self-reported attention, self-reported effort and completion time) was, however, weaker.

Results revealed that the self-reported attention and self-reported effort items were positively related to the mIERda. This suggests that participants reporting lower levels of effort and attention were more likely of being categorized as IER at the start and the end of the survey by the mIERda. Similarly, completion time was found to be negatively related to the mIERda throughout the survey. Such results suggest that participants who completed the survey more quickly were also more likely to engage in IER at any point in the survey. In other words, participants who took more time completing the survey were more likely of being attentive. Nonetheless, the strength of the correlation observed between self-reported attention, self-

reported effort, and survey completion time suggests that the mIERda did not share as much variance with these IER indicators as it did with attention checks and self-reported data inclusion.

Taken together, this suggests that the mIERda might be more strongly related with attention checks and data inclusion items, than the other indicators of IER. Such a result is not necessarily surprising if we examine *how* each of these indicators captures IER. For instance, in order for self-reported items to be an accurate measure of IER, participants need to be honest enough to declare that they were not fully attentive, and that they did not respond to the survey in an effortful manner. The accuracy of such measures is also dependent on participants paying attention to these specific items once they reach the end of the survey. If we make an example from the self-reported attention item to which there are five possible answers ("*0 - my full*"; "*4 - almost no*"), an IER participant who happens to be providing completely random answers have an even probability of picking any of the possible answers. Accordingly, while there is a 20% probability that they would accurately "classify" themselves by indicating that they have paid almost no attention to the survey, there is also a 20% probability of them responding that they provided *their full* attention to their survey. Contrastingly, an attentive participant who has given their full attention to the survey would have a 100% probability of accurately classifying themselves and providing the latter answer. Such patterns in responses could explain the small correlation observed between the mIERda, self-reported attention and self-reported effort items. It is plausible that the reliable responses provided by attentive participants might drive these items' correlations with the mIERda in the right direction even if IER participants might introduce error, which may limit their strength. Such hypotheses could also help us understand why the correlation between the mIERda and the self-reported data inclusion item seemed to be

stronger than with other self-reported items. In contrast with other self-reported measures of IER, there are two answer choices (Yes, No) to the self-reported data inclusion item. If we apply a similar reasoning, this means that IER participants now have a 50% probability of accurately classifying themselves without paying attention. Such a shift in hypothetical baseline accuracy could explain why the self-reported data inclusion item correlates more strongly with the mIERda than other items. The same principle applies to attention checks. Statistically-speaking, assuming a 5-point Likert scale and an attention check to which there is solely one correct answer, an IER participant has an 80% probability of correctly classifying themselves as IER (i.e., choosing an incorrect answer). This might explain why attention checks seemed to correlate more strongly with the mIERda than self-reported attention and self-reported effort items. Lastly, the weak correlation between the mIERda and survey completion time is not unexpected. In fact, such finding is coherent with past literature reporting a similar association of completion time with other measures of IER (Meade & Craig, 2012). Overall, the investigated indicators of IER were related, as expected, with the mIERda at the start and the end of the survey, which supports the convergent validity of the mIERda.

Discriminant Validity

The discriminant validity of the mIERda was supported by the non-significant correlations between the mIERda and social desirability. This result aligns with the nonsignificant or weak association between IER and social desirability reported in previous research (Meade & Craig, 2012). The mIERda was also uncorrelated with global controlled motivation and controlled motivation towards research participation. Such a finding is in line with results from previous research reporting a null correlation between controlled forms of motivation and IER (Cheung, 2019). This also suggests that controlled motivation, at a research

and global level, is not associated with engagement in IER. Although more research is needed to ascertain such a conclusion, these findings suggest that targeting participants' level of controlled motivation might not impact participants' engagement in IER and that other forms of motivation might represent more appropriate intervention targets.

Predictive Validity

Results additionally provided evidence of the predictive validity of the mIERda as shown by its positive correlation with amotivation at a global and research level, as well as the negative association of the mIERda with autonomous motivation at a global and research level. These results align with past research that has reported a positive association between IER and amotivation at various levels, that is research, academic, and global, as well as a negative association between IER and autonomous motivation at these levels (Cheung, 2019). The predictive validity of the mIERda was further validated by the negative correlation observed between conscientiousness and the mIERda throughout the survey. This finding aligns with results from previous studies reporting a negative association of conscientiousness with IER (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017). While this is beyond the scope of the present study, the positive association between amotivation, at a global and research level, and the mIERda (i.e., at the start and the end of the survey) suggests that researchers interested in preventing participants' engagement in IER might benefit from designing interventions which aim to decrease participants' level of amotivation. Likewise, the negative association between the mIERda (i.e., at the start and the end of the survey) and autonomous motivation, at a global and research level, suggests that researchers might benefit from fostering greater levels of autonomous motivation in participants. This could be achieved through autonomy supportive interventions such as by communicating more information about the importance of the study to

the participants, explaining to them the significance of their contribution, emphasizing their choice to complete or not the study, validating participants' emotional experience towards the survey, and minimizing the use of controlling techniques (e.g., Jang, 2008; Joussemet et al., 2008; Reeve et al., 2002).

Concurrent Validity

Results also provided evidence of the concurrent validity of the mIERda, as demonstrated by the positive correlation observed between the time of the semester and the IER status, suggesting that IER was more prevalent later in the semester. To our surprise, the correlation between the mIERda and the survey length was nonsignificant at the start of the survey and negative towards the end of the survey. Although participating in a shorter or longer survey was not associated with a different probability of IER at the start of the survey, participants who completed a longer survey seemed less likely to engage in IER at the end of the survey than those who completed a shorter survey. Survey length shared a similar association with other indicators of IER (i.e., completion time, attention checks, self-reported items) considered in the present study. The lack of association between engagement in IER at the start of the survey and survey length is not completely surprising if we consider past literature stating that engagement in IER increases as participants get exposed to more items (Berry et al., 1992). However, the negative association between survey length and engagement in IER at the end of the survey was opposed to our hypothesis and contradicts results arising from past literature suggesting that IER is more prevalence in longer surveys (Bowling et al., 2021, 2022; Gibson & Bowling, 2020). However, the fact that this effect extends to other indicators of IER suggests that the present results should not necessarily be interpreted as a lack of discriminant validity. Such results also

demonstrate how determinants of IER can act differently throughout the survey, which sheds light on the relevance of examining them at different instances throughout the survey.

While the present research does not allow us to understand why shorter surveys were associated with more IER at the end of the survey, we can hypothesize that participants who willingly chose to complete longer surveys also had more time at hand and were thus more willing to respond attentively to the survey. In opposition, it is also likely that participants who are more prone to engage in IER are more likely to complete shorter surveys in order to complete their research component as quickly as possible. More research is needed to examine this phenomenon as well as the role played by the time of the semester. Considering these variables, while accounting for participants' characteristics, would be of great relevance to understanding the circumstances under which participants might be more or less likely to engage in IER.

Known-Groups Validity

Lastly, the results from this study provided support for the known-groups validity of the mIERda. Our findings also suggested that men were more likely than women to be flagged as IER at any point throughout the survey. This result aligns with previous research reporting that the prevalence of IER is greater in men than in women (Braitman et al., 2025; Cheung, 2019; Schneider et al., 2018; Witt et al., 2011).

The mIERda Throughout the Survey

The results provided evidence of the construct validity of the mIERda at the start and the end of the survey. Given the dynamic nature of IER (Bowling et al., 2021), establishing the validity of the mIERda throughout the survey was deemed relevant to ensure the robustness of the results. The results suggest that the nature of their association of the mIERda with external criteria, except for survey length, remained similar irrespective of the moment of the survey.

Such results demonstrate how determinants of IER can act differently throughout the survey, which sheds light on the relevance of considering the factors that may foster IER at different instances throughout the survey.

To our surprise, the results have revealed that although the prevalence of IER increases throughout the survey, the strength of this effect was fairly limited. In fact, the observed prevalence of IER increased from 32.5% to 35.7% at the end of the survey. This suggests that although IER becomes more prevalent towards the end of the survey, there seems to be a considerable proportion of participants that shifts in and out of attentiveness throughout the survey. This is further supported by the moderate association between IER status at the start of the survey and IER status at the end of the survey. While more research is needed to understand this phenomenon, this supports the overall idea that IER is a dynamic process and that although some participants might be more prone than others to shift into a state of IER, engagement in IER might not solely be an “end of survey” phenomenon. These findings further highlight the relevance of performing IER screening on different portions of the survey to capture these fluctuations in engagement. Since our results indicate that approximately 50% of participants were found to engage in IER at any point throughout the survey, a nuanced approach to IER screening might be especially warranted and might allow researchers to preserve an adequate sample size. Overall, the prevalence of IER throughout the survey aligns with past research on the topic (Francavilla et al., 2019; Meade & Craig, 2012), which further supports the validity of the mIERda.

The mIERda vs. Other Indicators of IER

Results from the sensitivity analyses presented in the present study showed that the mIERda correctly classified as IER the vast majority of participants who reported that their

responses should not be included in the analysis because of poor data quality. The mIERda also classified as IER most of the participants who failed one or more attention checks. Not only were most of these participants identified as IER by the mIERda, but as participants failed more attention checks, they were also more likely of getting flagged as IER by the mIERda. It is also worth noting that moderate-to-strong positive correlations were observed between the mIERda and attention checks, as well as with the self-reported data inclusion item. While we could have expected these correlations to be even stronger, we must bear in mind results from the sensitivity (i.e., overlap) analyses demonstrating that the mIERda identified as IER most of the participants flagged by these indicators, while concurrently accounting for the results suggesting that these indicators failed to flag most of the participants identified as IER by the mIERda. Considering that a correlation is merely the square root estimate of the overlap between two variables, such correlations make sense: if these indicators overlapped as strongly with the mIERda as the mIERda overlapped with these indicators, the observed correlations would be stronger.

Such findings have considerable implications. Not only these further support the convergent validity of the mIERda, but it also suggests that while the mIERda could possibly replace these indicators (i.e., attention checks and self-reported data inclusion), these indicators could not replace the mIERda. Such results are unsurprising as we consider past literature supporting the use of multiple indicators of IER due to the limited sample-wide sensitivity of each IER indicator (Curran, 2016; Meade & Craig, 2012). Furthermore, as attention checks and the self-reported data inclusion item must be added to the survey prior to data collection, these results demonstrate that a sufficient level of screening quality can be obtained even without their inclusion. Such results position the mIERda as a potentially useful tool to screen archival data that do not contain such items or in surveys where these measures were not included. Taken

together, these results demonstrate the advantage of relying on the mIERda rather than individual indicators of IER and testify to the quality of the screening which can be performed without relying on said indicators.

Limitations and Future Research

Despite the insight provided by the present study, some limitations remain. For example, while results provided evidence of the convergent, discriminant, predictive, concurrent, and known-groups validity of the mIERda across 18 distinct datasets, the present research solely relied on datasets collected through a student research subject recruitment pool. Future research on the validation of the mIERda would benefit from analyzing samples drawn from the general population and with greater age diversity. It is also worth noting that we relied on datasets arising from a limited number of research programs in the fields of academic motivation, education, physical activity, motivation and well-being, and environmental psychology. As such, the surveys did not fully cover the breadth of the field of psychology, never mind other disciplines relying on surveys. Future research would benefit from validating the mIERda in datasets arising from other research programs and from other disciplines relying on psychometric measures. Furthermore, it is worth mentioning that caution is warranted when interpreting the results of this study as the association between IER and the self-reported validation variables might be influenced by IER. More precisely, it is not impossible that the pre-screen data, from which the measures of motivation and personality were drawn, is contaminated by IER. Such a possibility means that the effects reported in the present study might be biased. Future research would consequently benefit from relying on a multiple-rater approach or on an analytical approach that can accommodate or account for this bias, such as Bayesian inference. Such an approach would improve the robustness of the present findings. Lastly, since the present research

sought to establish the validity of the mIERda, we did not examine the association of the various correlates of IER while accounting for their mutual effects on IER. Future research would benefit from formally examining the determinants of IER using the mIERda.

Conclusion

The present research provided evidence of the validity of the multidimensional IER detection approach (mIERda), a novel IER detection approach. In this study, the construct validity of the mIERda was assessed across 18 datasets. Results from the present research offer strong support to the convergent, discriminant, predictive, concurrent, and known-groups validity of the mIERda. Overall, the present study positions the mIERda as a valid, encompassing and promising tool to facilitate and improve the detection of IER in survey data.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D. *American Journal of Preventive Medicine*, *10*(2), 77–84. https://doi.org/10.1016/S0749-3797(18)30622-6

Andrews, P., & Meyer, R. G. (2003). Marlowe–Crowne Social Desirability Scale and short Form C: Forensic norms. *Journal of Clinical Psychology*, *59*(4), 483–492. https://doi.org/10.1002/jclp.10136

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*(6), 2489–2505. https://doi.org/10.3758/s13428-020-01401-8

Aviv, A. L., Zelenski, J. M., Rallo, L., & Larsen, R. J. (2002). Who comes when: personality differences in early and later participation in a university subject pool. *Personality and Individual Differences*, *33*(3), 487–496. https://doi.org/10.1016/S0191-8869(01)00199-4

Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology*, *61*(2), 226–244. https://doi.org/10.1037/0022-3514.61.2.226

Braitman, A. L., Petrey, A. M., Shipley, J. L., Ayala Guzman, R., Renzoni, E., Looby, A., & Bravo, A. J. (2025). Check your data before you wreck your model: The impact of careless

responding on substance use data quality. *Alcohol, Clinical and Experimental Research*, *49*(4), 941–951. https://doi.org/10.1111/acer.70024

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*, *4*(3), 340–345. https://doi.org/10.1037/1040-3590.4.3.340

Bowling, N. A., Gibson, A. M., Houpt, J. W., & Brower, C. K. (2021). Will the Questions Ever End? Person-Level Increases in Careless Responding During Questionnaire Completion. *Organizational Research Methods*, *24*(4), 718–738. https://doi.org/10.1177/1094428120947794

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who Cares and Who Is Careless? Insufficient Effort Responding as a Reflection of Respondent Personality. *Journal of Personality and Social Psychology*, *111*(2), 218–229. http://dx.doi.org/10.1037/pspp0000085.supp

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor–Davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress*, *20*(6), 1019–1028. https://doi.org/10.1002/jts.20271

Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A New Scale for Measuring Adults' Prosocialness. *European Journal of Psychological Assessment*, *21*(2), 77–89. https://doi.org/10.1027/1015-5759.21.2.77

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Carver, C. S. (1997). You want to measure coping but your protocol's too long: Consider the brief COPE. *International Journal of Behavioral Medicine*, *4*(1), 92–100. https://doi.org/10.1207/s15327558ijbm0401_6

Cassidy, S. (2016). The Academic Resilience Scale (ARS-30): A New Multidimensional Construct Measure. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01787

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., Duriez, B., Lens, W., Matos, L., Mouratidis, A., Ryan, R. M., Sheldon, K. M., Soenens, B., Van Petegem, S., & Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, *39*(2), 216–236. https://doi.org/10.1007/s11031-014-9450-1

Cheung, S. (2019). *Understanding Insufficient Effort Responding from a Self-Determination Theory Perspective* [Doctoral Thesis, University of Ottawa]. uO Research. https://ruor.uottawa.ca/items/5db7e905-2c10-42a6-a335-6bf73341c438

Cheval, B., Maltagliati, S., Courvoisier, D. S., Marcora, S., & Boisgontier, M. P. (2024). Development and validation of the physical effort scale (PES). *Psychology of Sport and Exercise*, *72*, 102607. https://doi.org/10.1016/j.psychsport.2024.102607

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385. https://doi.org/10.2307/2136404

Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International Physical

Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise*, *35*(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134. https://doi.org/10.1016/0092-6566(85)90023-6

Deci, E., & Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer New York. https://doi.org/10.1007/978-1-4899-2271-7

de Calvo, M. P. C., & Reich, D. A. (2007). Spontaneous correction in the behavioral confirmation process: The role of naturally-occurring variations in self-regulatory resources. *Basic and Applied Social Psychology*, *29*(4), 351–364. https://doi.org/10.1080/01973530701665132

Diener, E., Emmons, R. A., Larsem, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Ekkekakis, P., Zenko, Z., & Vazou, S. (2021). Do you find exercise pleasant or unpleasant? The Affective Exercise Experiences (AFFEXX) questionnaire. *Psychology of Sport and Exercise*, *55*, 101930. https://doi.org/10.1016/j.psychsport.2021.101930

Elliot, A. J., & Thrash, T. M. (2010). Approach and Avoidance Temperament as Basic Dimensions of Personality. *Journal of Personality*, *78*(3), 865–906. https://doi.org/10.1111/j.1467-6494.2010.00636.x

Fischer, D. G., & Fick, C. (1993). Measuring Social Desirability: Short Forms of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, *53*(2), 417–424. https://doi.org/10.1177/0013164493053002011

Fraley, R. C., Heffernan, M. E., Vicary, A. M., & Brumbaugh, C. C. (2011). The experiences in close relationships—Relationship Structures Questionnaire: A method for assessing attachment orientations across relationships. *Psychological Assessment*, *23*(3), 615–625. https://doi.org/10.1037/a0022898

Francavilla, N. M., Meade, A. W., & Young, A. L. (2019). Social Interaction and Internet-Based Surveys: Examining the Effects of Virtual and In-Person Proctors on Careless Response. *Applied Psychology*, *68*(2), 223–249. https://doi.org/10.1111/apps.12159

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102. https://doi.org/10.1186/1479-5868-9-102

Gauthier, A. J. (2025). *The multidimensional Insufficient Effort Responding detection approach (mIERda)*. https://github.com/AJGauthier/mIERda.

Gauthier, A. J., Pelletier, L. G., & Gravel, C. (2025). *Sorting the Good from the mIERda – Introducing and Testing the performance of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)* [Unpublished manuscript]. School of Psychology, University of Ottawa.

Gibson, A. M., & Bowling, N. A. (2020). The Effects of Questionnaire Length and Behavioral Consequences on Careless Responding. *European Journal of Psychological Assessment*, *36*(2), 410–420. https://doi.org/10.1027/1015-5759/a000526

Gratz, K. L., & Roemer, L. (2004). Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41–54. https://doi.org/10.1023/B:JOBA.0000007455.08539.94

Grimby, G., Börjesson, M., Jonsdottir, I. H., Schnohr, P., Thelle, D. S., & Saltin, B. (2015). The "Saltin–Grimby Physical Activity Level Scale" and its application to health research. *Scandinavian Journal of Medicine & Science in Sports*, *25*(S4), 119–125. https://doi.org/10.1111/sms.12611

Guertin, C., Barbeau, K., Pelletier, L., & Martinelli, G. (2017). Why do women engage in fat talk? Examining fat talk using Self-Determination Theory as an explanatory framework. *Body Image*, *20*, 7–15. https://doi.org/10.1016/j.bodyim.2016.10.008

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2025). Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: Factors influencing careless responding. *Psychological Methods*, *30*(2), 374–392. https://doi.org/10.1037/met0000580

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and Deterring Insufficient Effort Responding to Surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

Jang, H. (2008). Supporting students' motivation, engagement, and learning during an uninteresting activity. *Journal of Educational Psychology, 100*(4), 798–811. https://doi.org/10.1037/a0012841

Joussemet, M., Landry, R., & Koestner, R. (2008). A self-determination theory perspective on parenting. *Canadian Psychology / Psychologie canadienne, 49*(3), 194–200. https://doi.org/10.1037/a0012754

Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, *10*(JUN). https://doi.org/10.3389/fpsyg.2019.01258

Kam, C. C. S., & Meyer, J. P. (2015). How Careless Responding and Acquiescence Response Bias Can Influence Construct Dimensionality: The Case of Job Satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Kim, D., Lee, Y., Lee, J., Nam, J. K., & Chung, Y. (2014). Development of Korean Smartphone Addiction Proneness Scale for Youth. *PLoS ONE*, *9*(5), e97920. https://doi.org/10.1371/journal.pone.0097920

Kwon, M., Kim, D.-J., Cho, H., & Yang, S. (2013). The Smartphone Addiction Scale: Development and Validation of a Short Version for Adolescents. *PLoS ONE*, *8*(12), e83558. https://doi.org/10.1371/journal.pone.0083558

Lavergne, K. J., & Pelletier, L. G. (2015). Predicting individual differences in the choice of strategy to compensate for attitude-behaviour inconsistencies in the environmental domain. *Journal of Environmental Psychology*, *44*, 135–148. https://doi.org/10.1016/j.jenvp.2015.10.001

Malach-Pines, A. (2005). The Burnout Measure, Short Version. *International Journal of Stress Management*, *12*(1), 78–88. https://doi.org/10.1037/1072-5245.12.1.78

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*(1), 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Martin, A. J. (2013). Academic buoyancy and academic resilience: Exploring 'everyday' and 'classic' resilience in the face of academic adversity. *School Psychology International*, *34*(5), 488–500. https://doi.org/10.1177/0143034312472759

Mazza, N., & Arthur, W., Jr. (2025). Correlates of Careless Responding: Trait and State Antecedents and Criteria. *Human Performance*, *38*(5), 243–264. https://doi.org/10.1080/08959285.2025.2552213

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Meng, H., Cao, H., Hao, R., Zhou, N., Liang, Y., Wu, L., Jiang, L., Ma, R., Li, B., Deng, L., Lin, Z., Lin, X., & Zhang, J. (2020). Smartphone use motivation and problematic smartphone use in a national representative sample of Chinese adolescents: The mediating roles of smartphone use time for various activities. *Journal of Behavioral Addictions*, *9*(1), 163–174. https://doi.org/10.1556/2006.2020.00004

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom. *Theory and Research in Education*, *7*(2), 133–144. https://doi.org/10.1177/1477878509104318

Ntoumanis, N., Ng, J. Y. Y., Prestwich, A., Quested, E., Hancox, J. E., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Lonsdale, C., & Williams, G. C. (2021). A meta-analysis of self-

determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. *Health Psychology Review*, *15*(2), 214–244. https://doi.org/10.1080/17437199.2020.1718529

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Pelletier, L. G., Dion, S. C., Slovinec-D'Angelo, M., & Reid, R. (2004). Why do you regulate what you eat? Relationships between forms of regulation, eating behaviors, sustained dietary behavior change, and psychological adjustment. *Motivation and Emotion*, *28*(3), 245–277. https://doi.org/10.1023/B:MOEM.0000040154.40922.14

Pelletier, L. G., Dion, S., Tuson, K., & Green-Demers, I. (1999). Why Do People Fail to Adopt Environmental Protective Behaviors? Toward a Taxonomy of Environmental Amotivation. *Journal of Applied Social Psychology*, *29*(12), 2481–2504. https://doi.org/10.1111/j.1559-1816.1999.tb00122.x

Pelletier, L. G., Legault, L. R., & Tuson, K. M. (1996). The Environmental Satisfaction Scale. *Environment and Behavior*, *28*(1), 5–26. https://doi.org/10.1177/0013916596281001

Pelletier, L. G., Tuson, K. M., Green-Demers, I., Noels, K., & Beaton, A. M. (1998). Why are you doing things for the environment? The Motivation Toward the Environment Scale (MTES). *Journal of Applied Social Psychology*, *28*(5), 437–468. https://doi.org/10.1111/j.1559-1816.1998.tb01714.x

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10 item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Reeve, J., Jang, H., Hardre, P., & Omura, M. (2002). Providing a rationale in an autonomy-supportive way as a strategy to motivate others during an uninteresting activity. *Motivation and Emotion, 26*(3), 183–207. https://doi.org/10.1023/A:1021711629417

Rocchi, M. A., Wilson, P. M., Sylvester, B. D., Guerin, E., Shi, Z., & Sweet, S. N. (2023). Toward brief tools assessing motivation for exercise: Rationale and development of a six- and 12-item version of the Behavioral Regulation in Exercise 3 (BREQ3) Questionnaire. *Sport, Exercise, and Performance Psychology*, *12*(3), 205–227. https://doi.org/10.1037/spy0000321

Rocchi, M., Pelletier, L., Cheung, S., Baxter, D., & Beaudry, S. (2017). Assessing need-supportive and need-thwarting interpersonal behaviours: The Interpersonal Behaviours Questionnaire (IBQ). *Personality and Individual Differences*, *104*, 423–433. https://doi.org/10.1016/j.paid.2016.08.034

Rosenberg, M. (1965). Rosenberg Self-Esteem Scale. In *PsycTESTS Dataset*. https://doi.org/10.1037/t01038-000

Ryan, R. M. (Ed.). (2023). *The Oxford handbook of self-determination theory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197600047.001.0001

Ryan, R. M., & Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. The Guilford Press. https://doi.org/10.1521/978.14625/28806

Ryan, R. M., & Frederick, C. (1997). On Energy, Personality, and Health: Subjective Vitality as a Dynamic Reflection of Well-Being. *Journal of Personality*, *65*(3), 529–565. https://doi.org/10.1111/j.1467-6494.1997.tb00326.x

Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research*, *27*(4), 1077–1088. https://doi.org/10.1007/s11136-017-1767-2

Sheu, H.-B., Chong, S. S., Chen, H.-F., & Lin, W.C. (2014). Well-being of Taiwanese and Singaporean college students: Cross-cultural validity of a modified social cognitive model. *Journal of Counseling Psychology*, *61*(3), 447–460. https://doi.org/10.1037/cou0000018

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, *15*(3), 194–200. https://doi.org/10.1080/10705500802222972

Soldatos, C. R., Dikeos, D. G., & Paparrigopoulos, T. J. (2000). Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. *Journal of Psychosomatic Research*, *48*(6), 555–560. https://doi.org/10.1016/S0022-3999(00)00095-7

Spence, G. B., & Oades, L. G. (2011). Coaching with self-determination in mind: Using theory to advance evidence-based coaching practice. *International Journal of Evidence Based Coaching and Mentoring*, *9*(2), 37–55.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder. *Archives of Internal Medicine*, *166*(10), 1092. https://doi.org/10.1001/archinte.166.10.1092

Spreng, R. N., McKinnon*, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to Multiple Empathy Measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Tuckman, B. W. (1991). The Development and Concurrent Validity of the Procrastination Scale. *Educational and Psychological Measurement*, *51*(2), 473–480. https://doi.org/10.1177/0013164491512022

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C. B., & Vallières, E. F. (1992). Academic Motivation Scale. *Educational and Psychological Measurement*, *52*. https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft25718-000

Ward, MK., & Meade, A. W. (2023). Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices. *Annual Review of Psychology*, *74*(1). https://doi.org/10.1146/annurev-psych-040422-045007

Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417–430. https://doi.org/10.1016/J.CHB.2017.06.032

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Williams, G. C., Grow, V. M., Freedman, Z. R., Ryan, R. M., & Deci, E. L. (1996). Motivational predictors of weight loss and weight-loss maintenance. *Journal of Personality and Social Psychology*, *70*(1), 115–126. https://doi.org/10.1037/0022-3514.70.1.115

Witt, E. A., Donnellan, M. B., & Orlando, M. J. (2011). Timing and selection effects within a psychology subject pool: Personality and sex matter. *Personality and Individual Differences*, *50*(3), 355–359. https://doi.org/10.1016/J.PAID.2010.10.019

Woods, C. M. (2006). Careless Responding to Reverse-Worded Items: Implications for Confirmatory Factor Analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3). https://doi.org/10.1007/s10862-005-9004-7

Table 3.1

Aggregated Socio-Demographic Information

Variables	*n*	%
Gender		
Women	3196	63.7
Men	1279	25.5
Other	332	6.6
Undisclosed	207	4.1
Age		
16-19 years	3369	67.2
20-25 years	1443	28.8
26 years and older	90	1.8
Undisclosed	112	2.2
Ethnicity		
Arab/Arab-African	553	11.0
Asian	1018	20.3
Black	819	16.3
Hispanic	87	1.7
Indigenous	31	0.6
Mixed Heritage	183	3.6
White	2104	42.0
Other	139	2.8
Undisclosed	80	1.6

Note. $N = 5014$. Due to the limited sample size of some datasets, sociodemographic details were reported at a research program level to ensure the anonymity of participants. The values reported for *n* and % denotes the number and proportion of distinct respondents

Table 3.2

Summary of Constructs and Measures for the Validation of the mIERda

Constructs	Measures	Example	Source	Appears in research program				
				1	2	3	4	5
		Convergent Validity						
Insufficient Effort Responding	Attention checks (Huang et al., 2012; Meade & Craig, 2012), 3-5 items	*"Please respond 'Moderately Agree' to this item"* *"I eat cement occasionally"*	Questionnaire (3 to 5 items randomly positioned in the questionnaire)	X	X	X	X	X
	Self-reported attention 1 item, 0-4 scale	*"I gave this study ____ attention"*	Questionnaire (End)	X	X	X	X	X
	Self-reported effort (Meade & Craig, 2012), 1 item, 0-4 scale	*"I put forth____ effort towards this study"*	Questionnaire (End)	X	X	X	X	X
	Self-reported data quality (Meade & Craig, 2012), 1 item, Yes-No scale	"*In your honest opinion, should we use your data? You will be fully compensated for your participation no matter your answer"*	Questionnaire (End)	X	X	X	X	X
	Completion time (in seconds) from the start to the submission of the survey	N/A	Survey platform internal timer	X	X	X	X	X
		Discriminant Validity						
Social Desirability	Marlowe-Crowne Social Desirability Scale - Short Form (Fischer & Fick, 1993), 13 items, True-False scale	"*I never hesitate to go out of my way to help someone in trouble"*	Questionnaire	X		X	X	

Constructs	Measures	Example	Source	Appears in research program				
				1	2	3	4	5
Global Controlled Motivation	Controlled motivation subscale from the : Global Motivation Scale (Pelletier et al., 2004), 1-7 Likert scale, 6 items, 3 subscales	*"In general, I do things... in order to show others what I am capable of"*	Pre-screen	X	X	X	X	X
Controlled Motivation Towards Research	Controlled motivation subscale from the: Motivation Towards Research Scale (Cheung, 2019), 1-7 Likert scale, 6 items, 3 subscales	*"Why you participate in research... I would feel guilty if I did not"*	Pre-screen	X	X	X	X	X
		Predictive Validity						
Global Autonomous Motivation	Autonomous motivation subscale from the: Global Motivation Scale (Pelletier et al., 2004), 1-7 Likert scale, 6 items, 3 subscales	*"In general, I do things... because they reflect what I value the most in life"*	Pre-screen	X	X	X	X	X
Autonomous Motivation Towards Research	Autonomous motivation subscale from the: Motivation Towards Research Scale (Cheung, 2019), 1-7 Likert scale, 6 items, 3 subscales	*"Why you participate in research... Participating in research is interesting"*	Pre-screen	X	X	X	X	X
Global Amotivation	Amotivation subscale from the: Global Motivation Scale (Pelletier et al., 2004), 1-7 Likert scale, 6 items, 3 subscales	*"In general, I do things... because they reflect what I value the most in life"*	Pre-screen	X	X	X	X	X
Amotivation Towards Research	Amotivation subscale from the: Motivation Towards Research Scale (Cheung,	*"Why you participate in research... I would feel guilty if I did not"*	Pre-screen	X	X	X	X	X

Constructs	Measures	Example	Source	Appears in research program				
				1	2	3	4	5
	2019), 1-7 Likert scale, 6 items, 3 subscales							
Conscientiousness	Conscientiousness subscale from the: Big Five Inventory (Rammstedt & John, 2007), 1-5 Likert scale, 2 items.	*"I see myself as someone who does a thorough job"*	Pre-screen	X	X	X	X	X
	Concurrent Validity							
Time of the semester	Obtained by computing the difference in weeks between the pool opening date and the moment of participation	N/A	Questionnaire	X	X	X	X	X
Survey Length	Obtained by computing the total number of items within a survey.	N/A	Questionnaire	X	X	X	X	X
	Known-Groups Validity							
Gender	1 item measure of gender identity, 12 response choices.	"*Please indicate which of the following responses best fits with your gender identity.*"	Pre-screen	X	X	X	X	X

Note. Research programs: 1 = Academic Resilience (datasets id #13 - 19), 2 = Cellphone use in Classroom (dataset id #1); 3 = Motivation and Well-being (datasets id #9 - 12); 4 = Physical Activity (datasets id #2 - 8); 5 = Pro-Environmental Behaviours (datasets id #20 - 23)

Table 3.3

Aggregated Descriptive Statistics of the Numerical Validation Variables

Variables	*n*	*M*	*SD*	*Min*	*Max*	*Skew*	*Kurt*
Proportion failed Attention checks	5006	0.18	0.28	0.00	1.00	1.50	-1.34
Completion Time[1,2]	5014	0.00	1.00	-2.01	3.10	1.08	1.07
Social Desirability	4243	5.06	1.97	0.00	10.00	-0.14	-0.31
Global Autonomous Motivation	4992	4.88	1.15	1.00	7.00	-0.44	0.21
Global Controlled Motivation	4992	3.83	1.39	1.00	7.00	-0.06	-0.46
Global Amotivation	4949	2.81	1.54	1.00	7.00	0.54	-0.45
Research Autonomous Motivation	5000	4.86	1.32	1.00	7.00	-0.29	-0.29
Research Controlled Motivation	5000	4.66	1.24	1.00	7.00	-0.24	0.37
Research Amotivation	4978	1.84	1.19	1.00	7.00	1.76	3.36
Conscientiousness	5000	3.62	0.84	1.00	5.00	-0.16	-0.44
Survey Length	5014	183.64	43.77	108	299	0.90	1.92
Time of Semester	4937	7.27	3.85	0.00	12.51	-0.35	-1.27

Note. $p < .05$. [1]Standardized variable. [2]Log transformed variable.

Table 3.4

Descriptive Statistics of the Categorical Validation Variables

Variables	*n*	%
Attention checks		
0 failed attention checks	3342	66.7
1 failed attention checks	929	18.5
2 failed attention checks	448	8.9
≥ 3 failed attention checks	295	5.9
Self-reported effort ("I put forth ____ effort towards this study")		
a lot of (0)	1381	35.7
quite a lot of (1)	1814	46.9
some (2)	524	13.6
very little (3)	80	2.1
almost no (4)	66	1.7
Self-reported attention ("I gave this study ____ attention")		
my full (0)	1991	47.7
very little of (1)	1647	39.4
some of my (2)	388	9.3
most of my (3)	101	2.4
my full (4)	51	1.2
Self-reported data inclusion ("*In your honest opinion, should we use your data? You will be fully compensated for your participation no matter your answer.")*		
Yes (0)	3762	89.5
No (1)	441	10.5

Note. Self-Reported data inclusion: 0 = keep data, 1 = exclude data.

Table 3.5

Pooled Correlation Estimates between the Multidimensional Insufficient Effort Responding Detection Approach Classification at the Start and End of the Survey and Validation Variables.

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.47* [.45, .49]	
Convergent Validity		
Attention checks	.52* [.51, .53]	.44* [.43, .45]
Self-reported effort	.19* [.16, .22]	.19* [.16, .22]
Self-reported attention	.23* [.20, .26]	.22* [.19, .24]
Self-reported data inclusion	.47* [.44, .49]	.46* [.43, .48]
Completion time[1,2]	-.15* [-.18, -.12]	-.12* [-.15, -.09]
Discriminant Validity		
Social desirability	-.03 [-.08, .01]	-.02 [-.06, .03]
Global controlled motivation	.00 [-.03, .02]	.01 [-.01, .04]
Research controlled motivation	-.01 [-.04, .02]	-.01 [-.04, .02]
Predictive Validity		
Global autonomous motivation	-.10* [-.13, -.07]	-.04* [-.06, -.01]
Global amotivation	.11* [.08, .14]	.10* [.08, .13]
Research autonomous motivation	-.05* [-.07, -.02]	-.05* [-.08, -.03]
Research amotivation	.10* [.08, .13]	.11* [.08, .13]
Conscientiousness	-.16* [-.18, -.13]	-.13* [-.15, -.10]
Concurrent Validity		
Time of semester[1]	.10* [.07, .13]	.07* [.04, .10]
Survey length[3]	-.01 [-.04, .02]	-.04* [-.06, -.01]
Known-Groups Validity		
Gender	.08* [.05, .10]	.06* [.03, .09]

Note. N = 3865-5014. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive; 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. [3]Since the variability in number of items was limited within each research programs, the correlation was computed at an aggregated level.

Table 3.6

Aggregated mIERda Overlap/Sensitivity with Attention Checks and Self-Reported Items

Variables	*n*	%
Number of failed attention checks		
≥ 1 Attention Check	1113	66.7
≥ 2 Attention Checks	582	78.3
≥ 3 Attention Checks	272	92.2
≥ 4 Attention Checks	43	100.0
5 Attention Checks	7	100.0
Self-Reported Data Inclusion		
No, don't use my data	358	81.2

Note. Values represent the number and proportion of participants flagged as IER by attention checks and self-reported items that were also flagged by the mIERda at any point throughout the survey.

Table 3.7

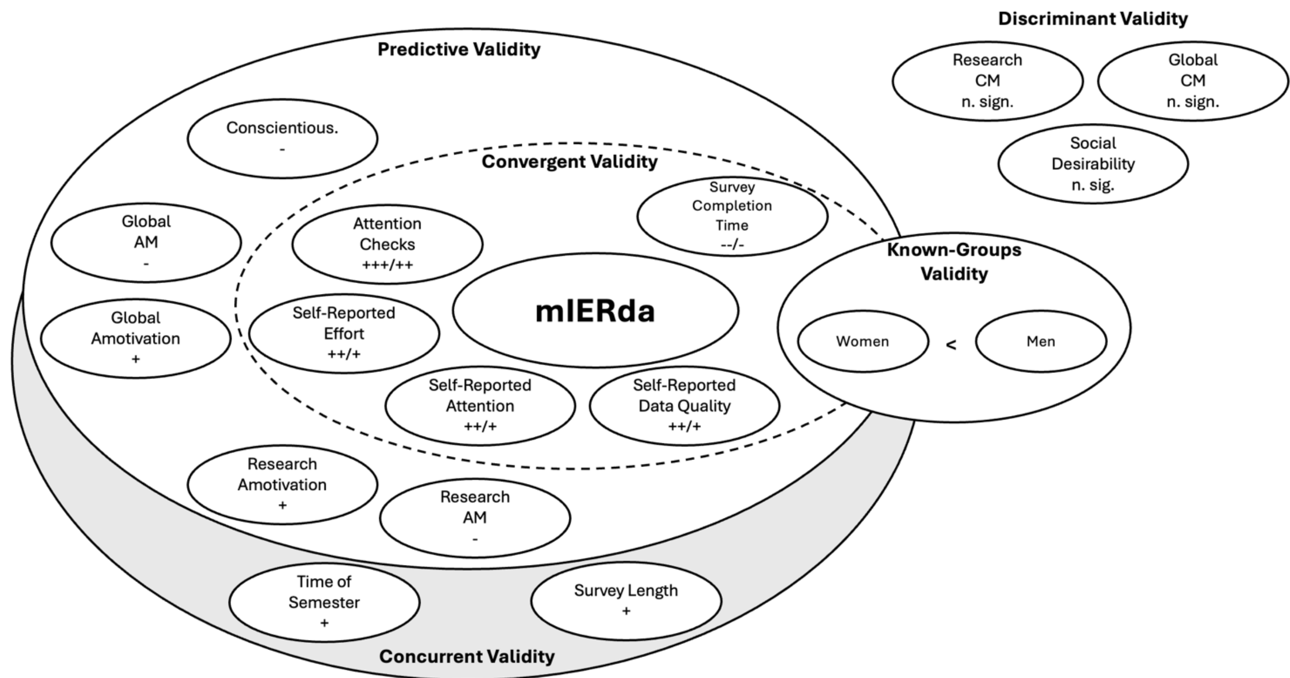
Prevalence of IER as per the Multidimensional Insufficient Effort Responding Approach by Datasets and by Research Programs

Dataset ID	mIERda			
	Start of survey		End of survey	
	n	%	*n*	%
	Academic Resilience			
13	86	40.2	67	31.3
17	7	24.1	8	27.6
18	8	14.8	20	37.0
19	26	24.3	38	35.5
Total	**127**	**31.4**	**133**	**32.9**
	Cellphone Use in the Classroom			
1	275	28.1	350	35.8
Total	**275**	**28.1**	**350**	**35.8**
	Motivation and Well-being			
9	64	32.5	67	34.0
10	106	40.0	88	32.8
11	38	27.7	36	26.3
12	144	30.8	184	39.3
Total	**352**	**32.9**	**375**	**35.0**
	Physical Activity			
2	122	29.9	122	29.9
3	21	50.0	27	64.3
5	11	39.3	11	39.3
7	75	36.7	98	48.5
8	86	29.4	88	30.0
Total	**315**	**32.4**	**344**	**35.4**
	Pro-Environmental Behaviours			
20	207	34.6	193	32.3
21	208	36.7	217	38.3
22	64	29.9	75	35.0
23	82	39.0	85	40.5
Total	**561**	**35.3**	**570**	**35.9**

Note. IER classification was established using the mIERda (0 = Attentive, 1 = IER)

Figure 3.1

Hypothesized Nomological Network of the multidimensional Insufficient Effort Responding detection approach (mIERda)



Note. --- : Strong negative correlation, -- : Moderate negative correlation, - : Weak negative correlation, +/- : Undetermined correlation, n. sig. : Non-significant correlation, + : weak positive correlation, ++ : Moderate positive correlation, +++ : Strong positive correlation. AM = Autonomous motivation; CM = Controlled motivation.

CHAPTER IV

Joint Exploration of the Contextual and Personal Determinants of Insufficient Effort Responding Over Time – A Self-Determination Theory Perspective

Ariane J. Gauthier[1] & Luc G. Pelletier[1]

Affiliation
[1]School of Psychology, University of Ottawa
136 Jean-Jacques-Lussier, Ottawa, Ontario, Canada K1N 9A8

Author's Contributions Statement CRediT

Ariane J. Gauthier : Conceptualization (lead), Methodology (lead), Validation, Formal Analysis, Data Curation, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing (lead), Visualization
Luc G. Pelletier : Conceptualization (supporting), Writing – Review & Editing (supporting)

Abstract

Given that results arising from past research has demonstrated the potential pervasive impact of insufficient effort responding (IER), researchers have been interested in understanding the determinants driving IER. Past research has highlighted that IER could be tied to various contextual (e.g., topic, survey length, time of the semester) and personal factors (e.g., motivation, personality, gender, age, ethnicity). However, studies examining these determinants often rely on cross-sectional designs and investigate these determinants individually. It thus remains unknown how well these determinants can predict IER over time, as well as across studies. In the present study, we rely on a prospective approach and analyze data from 23 datasets (N = 2216 participants, 5030 observations) arising from various research programs. Using a joint Bayesian multilevel logistic regression model, we investigated whether the role of contextual and personal determinants changed as respondents progress through the survey (i.e., at the start vs. end of the survey). Doing so, we also examined the role of motivation, at a domain level (i.e., motivation for research) and global level (i.e., overall motivational disposition) measured prior to participation in the different studies while accounting for other known determinants of IER. Results revealed a unique contribution of motivation which depended on the hierarchical level and evolved as participants progress through the survey. Higher levels of amotivation for research were associated with greater odds of engaging in IER throughout the survey, whereas higher global autonomous motivation predicted lesser odds of IER solely at the start of the survey. Results revealed an overall limited effect of contextual determinants, suggesting that IER is mostly related to respondents' personal characteristics. These findings are discussed through the lens of Self-Determination Theory, providing a useful and informative framework to

conceptualize meaningful interventions that could help reduce the prevalence of IER in future research.

Keywords: insufficient effort responding, careless responding, motivation, personality, survey research

Joint Exploration of the Contextual and Personal Determinants of Insufficient Effort Responding Over Time – A Self-Determination Theory Perspective

Once known as “random responding”, “content-independent responding”, “non-contingent responding”, “content non-responsivity”, or “careless responding”, the label insufficient effort responding (IER) was introduced by Huang and colleagues (2012) in an attempt to consolidate the literature on this topic. IER is defined as a response set in which the respondent answers a survey measure with low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses (Huang et al., 2012). Evidence suggests that IER respondents constitute between 3% and 50% of research samples (Francavilla et al., 2019; Meade & Craig, 2012). The presence of IER in research sample is concerning as it can have an insidious impact on scales’ factorial structure (Arias et al., 2020; Huang, Liu, & Bowling, 2015; Kam, 2019; Woods, 2006), lead to under- or over-estimation of internal consistency (Arias et al., 2020), distort the correlation between variables (Huang, Liu, & Bowling, 2015; Kam & Meyer, 2015), and diminish statistical power, construct validity and effect sizes (Kam & Meyer, 2015; Maniaci & Rogge, 2014; Oppenheimer et al., 2009).

Given that IER represents a significant threat to the integrity of data collected through online surveys, understanding the determinants of such behaviours is critical in order to prevent its occurrence. In order to understand what drives IER, past studies have been interested in understanding the role played by two broad categories of determinants: the contextual factors related to the research and respondents’ personal characteristics. The former refers to the contextual and environmental elements which can shape respondents experience when completing the survey. Survey length (Bowling et al., 2021; Gibson & Bowling, 2020; Ward et

al., 2017) and time of the semester (Aviv et al., 2002; de Calvo & Reich, 2007; Witt et al., 2011) are two commonly considered contextual factors. The latter corresponds to intra-individual differences which modulate how one might approach the survey such as personality (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017; Ward & Meade, 2023), gender, and age (Braitman et al., 2025; Grau et al., 2019; Schneider et al., 2018). Even if examining both categories of factors gives us a broad portrait of the factors which might promote the onset of IER, an important gap in the literature remains. Although highlighting the role of respondents' traits (e.g., personality) and characteristics (e.g., age and gender) might inform us as of *who* might be more likely to engage in IER, and highlighting contextual factors might inform as of *when* respondents might be more likely to engage in IER, these factors are hardly malleable. As such, investigating them does not provide us with relevant intervention targets which can be used to prevent IER. Therefore, there is a need to further investigate the contribution of a set of more intervention-sensitive factors, such as motivation. Interestingly, despite being at the very core of its definition (i.e., "low or little motivation"; Huang et al., 2012, p. 100), the association between motivation and IER remains vastly under-investigated.

This gap in the literature is problematic since understanding the motives underlying IER is imperative in order to design effective interventions which can enhance the responding effort deployed by respondents. Understanding how motivation relates to IER while accounting for known determinants of IER could also help researchers design their surveys in a way which promotes an effortful participation from respondents and enhance motivation. Furthermore, past studies have mostly relied on a cross-sectional design where IER and its determinants are simultaneously measured and solely assessed once. It thus remains unknown whether the association between IER and its various determinants is maintained when examining IER across

multiple studies and in contexts where the personal factors that are likely to affect IER (e.g., motivation) are assessed independently from IER in the target study. Additionally, past research has mainly focused on the determinants of IER by analyzing the survey as a whole (i.e., by measuring respondents' engagement in IER at a survey level). It therefore remains unknown whether the determinants that predict IER at the start of the survey differ from those that predict it at the end. The consideration of such distinction is especially relevant since IER is said to become more prevalent as respondents progress through the survey (e.g., Bowling et al., 2021) and since respondents' response effort may fluctuate over the course of the survey (Huang & DeSimone, 2021). Consequently, understanding the determinants of this shift in engagement could provide us with relevant insight to foster sustained attention as respondents progress through the survey. That is why the present research not only sought to shed light on the role of motivation while controlling for known determinants of IER, but also to investigate the role of these determinants throughout the survey.

Determinants of Insufficient Effort Responding

Several factors can influence the quality of one's response to an online survey. Over the past years, multiple studies have been dedicated to the exploration and examination of the contextual and personal determinants of IER. Below, we summarize the literature pertaining to both types of determinants and discuss their association with IER.

Contextual Determinants

Contextual determinants refer to all the external factors which shape the context in which the respondent completes the survey. More specifically, in university-based research subject recruitment pools, survey length, and time of the semester are two common factors which have been found to be associated with IER in past research.

Survey Length. A large body of research has already examined the relationship between survey length and IER (Bowling et al., 2021; Gibson & Bowling, 2020; Ward et al., 2017). Evidence suggests that longer surveys are associated with a higher prevalence of IER. More specifically, a study evaluating the prevalence of random responding (i.e., a type of IER) to a measure of personality suggests that although IER can be found from the start of a survey, where the respondent has been exposed to few items, it becomes even more prevalent toward the middle and end of longer surveys (Baer et al., 1997; Berry et al., 1992). This suggest that as participants were exposed to more items, they became more likely to engage in IER. Other studies have found that respondents assigned to a longer survey condition were more likely to engage in long string than respondents assigned to a shorter survey condition (Galesic & Bosnjak, 2009; Herzog & Bachman, 1981). Similarly, it was also found that respondents spent more time elaborating their response to a given item at the beginning of a survey than those who responded to the same item at the end of the survey (Galesic & Bosnjak, 2009) suggesting that respondents engagement waned as they were exposed to more items.

Taken together, the literature which as examined the role of survey length seem to suggest that respondents are more likely to engage in IER in longer survey, especially as they approach the end of the survey. This not only highlights the relevance of considering the role played by survey length on respondents' engagement in IER, but also of examining how this effect differs as participants progress within the survey (start. vs. end).

Time of Semester. In addition to survey length, the literature has identified the time of the semester as a predictor of IER in students recruited through university-based research subject recruitment pools. While such determinant is irrelevant for general population recruitment pools, past research found that the moment at which respondents chose to participate in research within

an academic semester might provide insight as it pertains to the quality of their responses. Some studies support that respondents partaking in research earlier in the semester (i.e., start of the semester) are less likely to engage in IER than respondents participating towards the end of the semester (Aviv et al., 2002; Cheung, 2019; Witt et al., 2011). It is posited that respondents' self-regulatory resources deplete throughout the semester, which might render them more likely to engage in IER once they reach the end of the semester (de Calvo & Reich, 2007). A study investigating the quality of research data in a university research subject recruitment pool found that respondents who participated in a study at the end of the semester reported providing lower quality data than respondents who participated at the start of the semester (Ebersole et al., 2016). Interestingly, evidence suggests that women and individuals with high levels of conscientiousness were not only more likely to participate in research earlier in the semester but were also less likely to engage in IER (Aviv et al., 2002; Witt et al., 2011). Globally, research suggests that the context in which the study takes place not only impacts the quality of the data, but that these factors might interact with personal determinants.

Personal Determinants

While research suggests that various contextual characteristics are associated with an increased prevalence of IER. The literature also provides evidence that multiple personal determinants are at play. In fact, past research reports age, ethnicity, and gender-related differences in odds of IER. Research also supports that personality, and motivation can impact one's propensity to engage in IER.

Age. The literature reports a mixed effect of age. While evidence remains scarce, one study reports that the probability that one would engage in IER decreases as respondent's age increases (Schneider et al., 2018). Another study, however, failed to find any significant effect of

age after controlling for the effect of respondents' personality, gender, education, and home country emancipation level (Grau et al., 2019). A thorough examination of the effect of age is thus warranted to better understand its association with IER.

Ethnicity. A recent study suggests that, after accounting for numerous other sociodemographic variables (e.g., gender, sexual orientation, and socioeconomic status), the propensity to engage in IER differed across respondents' ethnic identity (Braitman et al., 2025). Another study, which has specifically examined the cultural differences in IER, found that the likelihood to engage in IER differed depending on respondents' country of origin (Grau et al., 2019). Precisely, it was found that engagement in IER was negatively associated with respondents' country emancipation index. In other words, respondents from emancipated countries were more likely to respond attentively to the survey. Although more research is needed to understand the root cause of these differences, the authors suggest that culture might play a strong role on one's perception of the value of participating in research (Grau et al., 2019).

Gender. While more research is needed to explore the etiology of this phenomenon, evidence consistently supports that men are more prone to IER than women (Braitman et al., 2025; Cheung, 2019; Schneider et al., 2018). While the literature reports these gender differences, it is unlikely that one's gender is the "cause" of IER, this informs us that socialization (i.e., gender roles) might have a role to play in IER behaviours. In facts, research suggests that women tend to differ from men in terms of agreeableness, conscientiousness, emotional stability, extraversion, and openness to experience (Weisberg et al., 2011). Such gender difference in personality is relevant to consider given the known contribution of personality on IER. This not only highlights the relevance of considering gender, but also of accounting for the role of personality.

Personality. Due to the relative rank-order consistency of IER across time (Bowling et al., 2016), and the relative stability of personality traits across time and contexts, researchers have been interested in examining the role of personality on IER. Research has revealed that acquaintance- and self-reported conscientiousness, extraversion, agreeableness, and emotional stability are associated with a lower likelihood of IER (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017; Ward & Meade, 2023). This suggests that some individuals might be more likely than others to engage in IER. While these authors have argued that IER is likely to be an enduring interpersonal difference, it is worth mentioning that other studies have failed to find any significant association between personality and IER (Camus, 2015; Harms et al., 2017). Interestingly, Hasselhorn and colleagues (2025) failed to find any significant effect of personality after accounting for the role of motivation. Such results support the relevance of examining the role of personality while accounting for the effect of motivation.

Motivation. IER is depicted as a lack of motivation to respond and interpret survey items (Huang et al., 2012). Past research suggests that, in a university subject-pool setting, domain-specific (i.e., academic and research), and global levels of amotivation are positively associated with IER (Cheung, 2019; Gauthier & Pelletier, 2025). More recently, Hasselhorn and colleagues (2025) have explored the role of momentary motivation, in the context of a study examining the role of participants' motivation and mood on IER in ambulatory assessments. For seven days, respondents completed three to nine assessments per days. These assessments contained 19 to 29 items, and on each occasion, respondents were invited to report the extent to which they were motivated to respond to the survey items using a 1-item measure of motivation (i.e., "*I am currently motivated to answer the questions*"). After controlling for the effect of personality, it was found that respondents' average level of momentary amotivation negatively predicted IER,

but there was no effect of research motivation at a situational level. Similarly, Ashley and Shaughnessy (2021) have found a limited association between situational research intrinsic motivation and various indicators of IER (i.e., long string index, response time, psychometric synonyms). Overall, results suggest that motivation might be a determinant of IER but that it may be relevant to consider the hierarchical level of motivation (i.e., whether it reflects a general disposition that affects all behaviours, pertains to the domain of research, or relates to a specific occurrence of research participation). Fortunately, Self-Determination Theory offers an encompassing framework not only to understand why some individuals may engage in IER but also how this can be driven at various hierarchical levels.

Revisiting the Role of Motivation in IER: The Self-Determination Theory Perspective

Self-Determination Theory (Deci & Ryan, 1985) is a leading theory of human motivation. Used across a broad number of domains (Ryan, 2023), Self-Determination Theory is built on the principle that motivation can be defined in terms of quality, rather than in terms of quantity. By focusing on the quality of the motivation underlying behaviours, the theory offers a compelling theoretical framework to not only understand the reasons why behaviours are initiated, but also the factors promoting their maintenance. Over the last 40 years, this theory has been successfully applied in various domains, such as in education (e.g., Niemiec & Ryan, 2009), health behaviours (Ntoumanis et al., 2021), coaching (e.g., Spence & Oades, 2011), eating regulation (Guertin et al., 2017; Williams et al., 1996), and pro-environmental domain (Pelletier et al., 1998).

Self-Determination Theory proposes that there are six types of regulation that can be used to describe the reasons underpinning one's engagement in various behaviours. These types of regulation differ from one another in terms of their level of autonomy, or self-determination (i.e.,

motivational quality). The theory presents these regulations along a continuum of autonomy ranging from amotivation to intrinsic motivation.

Amotivation is the least self-determined form of motivation. It refers to the behaviours enacted without intentionality or meaningfulness. When experiencing amotivation, the person is disengaged and does the behaviour although they do not see the purpose or value of doing so (e.g., participating in research although it feels useless). Next on the continuum of self-determination, we find *controlled motivation*, which encompasses *external* and *introjected regulation*. *External regulation* refers to behaviours enacted as a response to factors or incentives external to the self, such as social pressure, to obtain a reward, or to avoid a punishment (e.g., participating in research to obtain a compensation). *Introjected regulation* defines behaviours carried out of shame, guilt, fear of exclusion and disapproval. Under introjected regulation, actions are ego-driven and performed to avoid such feelings (e.g., participating in research to avoid feeling guilty for not doing so). Lastly, there is *autonomous motivation*, which is the most self-determined form of motivation. Autonomous motivation, or self-determined motivation, encompasses identified, integrated, and intrinsic regulation. *Identified regulation* is the least self-determined form of autonomous motivation and refers to behaviours performed for a specific purpose and because they are deemed important to the person (e.g., participating in research is important). On the other hand, *integrated regulation* refers to behaviours enacted because are in accordance with the self. The behaviour can either allow the person to work towards a goal important to them or be in line with their core values (e.g., participating in research because doing so aligns with one's values). Finally, there is *intrinsic regulation* which underpins behaviours initiated out of sheer enjoyment and satisfaction (e.g., participating in research because one enjoys doing so). The theory posits that individuals with greater levels of

autonomous motivation also tend to remain more engaged, committed, and persistent when engaging in a task. (Deci & Ryan, 2017). Controlled motivation and amotivation do not confer these advantages. Individuals with higher levels of controlled motivation and amotivation are, in fact, more likely to become disengaged if external pressure is lifted or the task becomes inconvenient (Green-Demers, 1997). Theoretically, in a research participation context, this implies that individuals with greater levels of autonomous motivation should be less likely to engage in IER, whereas individuals guided by controlled or amotivated motives should be less likely to engage meaningfully with the survey which might translate into more IER. It is however worth noting that motivation can act at various levels. Although understanding how the quality of the motivation can influence one's engagement in a task, individuals operate and interact in various environments and contexts. As such, in order to understand the role played by motivation, we not only need to consider its degree of autonomy but also its level of specificity. Vallerand (1997), through his Hierarchical Model of Intrinsic and Extrinsic Motivation (HMIEM), proposes a framework to conduct such examination.

Hierarchical Model of Motivation

Vallerand's (1997) HMIEM states that motivation can be examined at three levels of specificity (i.e., global, domain-specific, and situational). The broader and least specific level, the *global level* of motivation, refers to general tendencies and one's personality. It reflects inter-individual differences in how one tends to regulate and approach behaviours through an autonomous, controlled, or amotivated orientation, irrespectively of life domains. Next, follows the *domain-specific level* of motivation. The domain-specific level is more specific than the global level. Rather than referring to general tendencies, the domain-specific level of motivation describes individuals' motivational orientations in a particular life domain, such as education,

community engagement, sports, or research participation. It describes one's motivational orientation for the tasks and behaviours tied to this domain. Motivation for behaviours of a specific domain might, however, change based on circumstances and vary across time. As such, the most specific level (i.e., *situational level*) of motivation refers to task-specific motivation within a given domain, such as an individual's motivation orientation at a specific point in time (e.g., when participating in a specific study or completing a given survey). These three levels of motivation are hierarchical, from global to situational.

According to the HMIEM, motivational orientation at a lower level of specificity (e.g., domain-specific) can influence motivation at a higher level (e.g., global), exerting a *bottom-up effect*. This posits, for instance, that participating in research out of genuine interest (i.e., autonomous domain-specific motivation) might translate into higher levels of autonomous motivation at a global level. Concurrently, it is posited that global motivation has a *top-down effect* and can influence domain-specific motivation which, in turn, can influence situational motivation (Vallerand, 1997). This suggests that individuals who are generally autonomously motivated in their life may also pursue more specific tasks under a similar orientation (Lavigne & Vallerand, 2010). In a research participation context, this means that respondents with greater autonomous motivation at a global level should be more likely be driven by autonomous motivation when engaging into domain specific tasks, such as participating in research. Accordingly, these respondents with greater levels of autonomous motivation, whether at a global or domain-specific level, are also expected to be more engaged each time they participate in research. Consequently, these participants should be less likely to become disengaged and should thus remain more attentive throughout the survey. Whereas individuals that are with greater levels of controlled motivation or amotivation, irrespectively of the hierarchical level,

should display lower level of engagement and be more likely to engage in IER when partaking in survey research. Additionally, the HMIEM postulates that more specific forms of motivation should be stronger predictors of one's behaviours than least specific forms of motivation. In other words, this implies that domain-specific motivation (i.e., research participation) should predict IER more strongly than motivation measured at a global level. This suggests that the effect of the various forms of motivation on respondents' engagement in IER should vary across hierarchical levels.

Motivational Orientation, Motivational Specificity, and IER

While evidence remains relatively scarce, research provides preliminary evidence that respondents' motivational orientation might impact the quality of responses provided by research respondents and that the effect of motivation differs across hierarchical levels. Below, we discuss the association between IER and the various forms of motivational orientation (i.e., amotivation, controlled motivation, and autonomous motivation) and review how these associations differ depending on the level of specificity.

Amotivation. Although Cheung (2019) reports that participants' levels of amotivation at various levels of specificity are associated with an increased likelihood of IER, they found that the effect of amotivation at a global level subsided after accounting for a more specific form of amotivation (i.e., academic motivation). Similarly, the effect of academic amotivation waned after accounting for an even more specific form of motivation (i.e., research participation). More recent research however suggests that increasing the specificity of the amotivation measure beyond a domain-specific level (i.e. research participation) might not improve our ability to predict IER. Indeed, Hasselhorn and colleagues (2025) have explored the role of amotivation in ambulatory assessments (i.e., repeated measurements of individuals in their daily lives over a set

period of time) and reported that amotivation for participating in research at a situational-level (i.e., current motivation for a specific assessment) was not associated with IER, but that the person-level of amotivation reported (i.e., average level of amotivation towards research) across assessments was positively associated with IER. This suggests that the effect of amotivation seem to depend on its hierarchical level (e.g., situational vs. domain-specific).

Controlled motivation. Although amotivation was found to be associated with more IER, the literature fails to report such association with controlled motivation. Results arising from Cheung (2019) indeed support that controlled motivation, regardless of its hierarchical level (i.e., global and domain-specific) might not be related to IER. This result was further corroborated by Hasselhorn and colleagues (2025) who have demonstrated that respondents who reported participating in the study for external motives (i.e., to obtain a compensation) were not more likely to engage in IER than others. Similarly, Gauthier and Pelletier (2025) reported a null association between IER and controlled motivation at various hierarchical levels (i.e., global and domain-specific). Taken together, these findings suggest that individuals with greater levels of controlled motives, at a global, domain-specific, or situational level do not seem more likely to engage in IER.

Autonomous motivation. Past research reports a negative association between IER and autonomous motivation (Cheung, 2019; Gauthier & Pelletier, 2025). As observed for amotivation, research supports that the effect of autonomous motivation on IER depended on its hierarchical level. More precisely, Cheung (2019) demonstrated that higher levels of autonomous motivation at a global level was associated with a decreased engagement in IER whereas more specific forms of motivation (i.e., academic and research) were not associated with IER. This differential effect is further exemplified by Ashley and Shaughnessy (2021) who reported that

situational intrinsic motivation (i.e., a form of situation autonomous motivation) was either uncorrelated or shared a weakly negative correlation with various indicators of IER (i.e., long string index, response time, psychometric synonyms). Taken together, such results suggest that autonomous motivation at a dispositional level (i.e., global) might be a stronger determinant of IER than more specific forms of motivation (i.e., situational or domain-specific).

Overall, these results suggest that respondents who generally feel a sense of choice and meaning in their life (i.e., who have higher levels of autonomous motivation) might be less prone to engaging in IER; whereas research respondents who generally feel like they do not have a sense of volition and purpose when participating in research (i.e., higher level of amotivation towards research) might be more prone to engage in IER. Although the past literature provides us with relevant insight to understand the association between IER, motivation, and other determinants of IER, limitations remain to be addressed.

Limitations of Past Research

Even if multiple studies have been dedicated to the exploration and examination of the determinants of IER and highlighted the role of personality (e.g., Bowling et al., 2016; Camus, 2015; Dunn et al., 2018; Harms et al., 2017; Ward et al., 2017; Ward & Meade, 2023), motivation (Ashley & Shaughnessy, 2021; Cheung, 2019; Hasselhorn et al., 2025), survey length (e.g., Bowling et al., 2021; Gibson & Bowling, 2020; Ward et al., 2017), time of the semester (Aviv et al., 2002; Cheung, 2019; Witt et al., 2011), age, ethnicity, and gender (e.g., Braitman et al., 2025; Grau et al., 2019; Schneider et al., 2018), multiple limitations remain.

In a first instance, while research examined the individual or joint effect of a few of these determinants, their effect has never been conjointly tested using data arising from various studies. Furthermore, although assessing the effect of these determinants using a cross-sectional

design can help us pinpoint the potential determinants of IER, results should be interpreted cautiously. Without controlling for respondent-level variance across multiple time points, establishing causality remains premature. More precisely, without a prospective or a longitudinal design where the determinants are measured across time, it remains unknown whether the positive association between IER and time of the semester (Aviv et al., 2002; Cheung, 2019; Witt et al., 2011) is tied to between- or within-person differences. In other words, it is unspecified whether this effect is tied to a selection bias where respondents who are more likely to engage in IER also tend to participate later in the semester (i.e., suggesting a between-person difference) or if this effect might be due to fatigue and that the simple fact of participating later in the semester leads to more IER (i.e., suggesting a within-person difference). In fact, the focus of the few studies reporting an examination of IER using a longitudinal approach has either been to examine the relative consistency of IER across time (Bowling et al., 2016) or the relationship of personal determinants (i.e., personality and momentary motivation) with IER (Hasselhorn et al., 2025). While such research provides us with insight as it pertains to the distinction between IER and attentive respondents in terms of personality and momentary motivation across time, it does not assess the impact of various contextual determinants of IER, such as study topics, survey length, and time of the semester. As such, it remains unknown whether these findings can fully be attributed to these personal determinants or if they are better explained by survey characteristics. This also implies that, to the best of our knowledge, the contribution of contextual determinants at a between- and within-person level remains completely unexplored.

Speaking of the survey's characteristics, the effect of study's topic and IER remains to be thoroughly investigated. While Hasselhorn and colleagues (2025) state that respondents' levels of interest in the study topic did not differ across attentive and IER respondents, their study

solely assessed respondents' interest for a single topic. In other words, they did not assess whether the prevalence of IER might vary across study topics. Furthermore, even if the study conducted by Hasselhorn and colleagues (2025) represents a first attempt to capture the effect of motivation on IER over time, it has important limitations. Firstly, Hasselhorn and colleagues' (2025) study measures motivation, during each assessment, using one item (i.e., "*I am currently motivated to answer the questions*") to which respondents responded using a 5-point Likert scale (i.e., "*1–Not at all*"; "*5–Very much*"). While this item can inform us on the "quantity" of motivation experienced by the respondent, it does not inform us about the "quality" of respondents' motivation as conceptualized by Self-Determination Theory (Deci & Ryan, 1985; Ryan & Deci, 2017). Such item, which targets the presence (or lack of motivation), could at best be interpreted as a measure of amotivation. Which, as defined by SDT, describes actions enacted without a sense of purpose or meaningfulness. While Hasselhorn and colleagues' (2025) study gives us a glimpse of the association between amotivation levels and IER status, the effect of autonomous and controlled motivations on behaviours of IER across time remains unknown. Such gap is problematic since IER might be guided by a dual process: on one hand, autonomous motivation could be related to sustained effort and good survey-taking etiquette (lower IER), while amotivation could be related to lack of effort and problematic survey-taking behaviours (higher IER). This, taken in conjunction with the hierarchical model of motivation previously introduced, warrants the exploration of different types of motivation in the prediction of IER across time. Furthermore, Hasselhorn and colleagues' (2025) did not assess motivation independently from the survey questions which limits the causal strength of their findings. Similarly, although the effect of motivation (i.e., amotivation, controlled motivation, autonomous motivation) across various levels of specificity, as defined by SDT, has been explored by Cheung

(2019), further research is needed to better understand this relationship. On one hand, past research failed to consider the role of motivation in conjunction with other determinants of IER such as the study characteristics (i.e., number of items, study topics), and other determinants such as personality traits. Consequently, it is still unknown whether IER is driven by domain-specific and global motivation, or perhaps the lack thereof, especially after accounting for these other determinants. On the other hand, Cheung (2019) relied on a cross-sectional design. This implies that the role of motivation at a global and domain-specific level on engagement in IER across time remains to be examined. The generalizability of past research is thus limited, highlighting the need to examine such questions in a more encompassing and general model.

Lastly, although past research has suggested that engagement in IER might differ at the start vs. the end of a survey (Galesic & Bosnjak, 2009), it remains unknown whether the determinants of that behavior also differ at these two time points. Past research has relied on measures of IER at a survey-level or focus on the role of survey length. Consequently, the role played by each determinant throughout the survey remains unclear. More specifically, it is unknown whether the determinants which fosters IER at the start of the survey differ from the determinants fostering IER at the end of the survey. For example, it is possible that autonomous motivation at a global level might impact how respondents approach the survey in general and acts as a protective factor at the start of the survey. In contrast, autonomous motivation at a domain-specific level (i.e., research participation) might have a stronger effect as the task increases in specificity (i.e., at the end of the survey). Such a gap in knowledge is problematic since understanding how determinants shape respondents' engagement as they progress through a survey is crucial to design targeted and relevant interventions.

Present Research

The present research sought to tackle the aforementioned limitations by investigating the joint contribution of known contextual and personal determinants of IER across studies of varying topics. More precisely, we aimed to examine the joint effect of domain-specific motivation, global motivation, personality, study length, study topics, and time of the semester on IER throughout the survey (i.e., start vs. the end of the survey), while controlling for various covariates (i.e., gender, age, and ethnicity) using a panel-model approach and a prospective design (i.e., personal determinants are measured prior from the participation in the studies and independently from IER). To achieve this goal, we performed IER screening on 23 datasets/studies arising from five distinct research programs and collected over the course of 32 months. We propose to examine the effect of these determinants on the likelihood of IER as respondents progress through the survey using a joint Bayesian multilevel logistic regression model. This analytical approach offers a more robust examination the determinants of IER not only by shedding light on the main determinants of IER but also by allowing us to account for the variance attributable to respondents. It also allows use to disaggregate the between-person and within-person effects of contextual determinants of IER.

Hypotheses

It was hypothesized that the study topic would influence one's likelihood to engage in IER (1a; within-person effect) and that some research programs would attract participants who are more likely to engage in IER (1b; between-person effect). Furthermore, we posit the hypothesis that respondents who, on average, completed longer surveys would be more likely to engage in IER (2a). In accordance with the literature supporting that the likelihood of IER is greater in longer surveys (Bowling et al., 2021; Gibson & Bowling, 2020; Ward et al., 2017), it

was hypothesized that respondents completing longer surveys than usual (compared to their own average) would be more likely to engage in IER at the end of the survey (2b). In line with the literature supporting that the prevalence of IER is greater towards the end of the semester (Aviv et al., 2002; Cheung, 2019; Witt et al., 2011), it was also hypothesized that the likelihood of IER would be greater in respondents who tend to complete their research participation component later in the semester (3a) and when respondents complete their research participation component later than usual (3b). Building on past research, it was hypothesized that personality would be a meaningful predictor of IER. More specifically, in accordance with the literature, it was hypothesized that respondents' openness to experience (4a), conscientiousness (4b), extraversion (4c), and agreeableness (4d) would be associated with a decreased likelihood of IER; whereas neuroticism (4e) would be associated with an increased likelihood of IER (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017; Ward & Meade, 2023). With regards to motivation, it was hypothesized that, after accounting for the role of contextual and other personal determinants of IER, motivation for research participation and motivation at a global level would influence the likelihood of IER. More precisely, building on Self-Determination Theory (Deci & Ryan, 2017), it was hypothesized that research amotivation (5a) and global amotivation (6a) would be positive predictors of IER. However, based on findings from Cheung (2019) and Gauthier & Pelletier (2025), it was hypothesized that controlled motivation at the research (5b) and global levels (6b) would not be predictors of IER. Also, it was hypothesized that research (5c) and global (6c) autonomous motivation would predict a decreased likelihood of engaging in IER. Finally, in accordance with Vallerand's Hierarchical Model of Motivation (1996), it was posited that the predictive effect of motivation for research participation would be stronger than the effect of

global motivation (7) as motivation for research is more closely related to the dependant variables investigated, that is IER.

Method

Data from this study arise from a larger research project. The participants included in this study are the ones who appeared in multiple occasions in the dataset (i.e., completed two or more studies).

Participants

The sample included a total of 2216 respondents who appeared in two or more datasets. More specifically, 1838 respondents appeared in two datasets, 200 appeared in three datasets, 149 appeared in four datasets, 17 appeared in five datasets, 11 respondents appeared in six datasets and one respondent appeared in seven datasets. The median age at pre-screen (T_0) completion was 19 years old (*n* = 1079, range= 16 – 31+). Most respondents 65.9% identified as women (*n* = 1462), while 22.4% identified as men (*n* = 497), 5.4% identified either as bi-spiritual, two-spirit, non-binary, gender neutral, agender, gender fluid, gender queer, or otherwise (*n* = 120), and 6.2% did not disclose their gender (*n* = 137). Most respondents identified as white (*n* = 956; 43.1%), Asian (*n* = 365; 16.5%), Black (*n* = 332; 15.0%), or Arab (*n* = 301, 13.6%). The remainder of the sample identified as indigenous (*n* = 19, 0.86%), Hispanic (*n* = 32, 1.4%), of mixed heritage (*n* = 92, 4.2%), otherwise (*n* = 56, 2.5%) or failed to disclose their ethnicity (*n* = 63, 2.8%). For an overview of the aggregated sociodemographic statistics refer to Table 4.1. For an overview of the descriptive statistics by study topic, refer to Appendix E (*Table E-4.3*).

Procedure

To examine the determinants of IER, respondents were recruited from September 2021 to April 2024, through an eastern Canadian university research subject recruitment pool.

Respondents were enrolled in introductory or second-year courses with a research participation component in psychology, business management, communications, or linguistics. As they enrolled to the recruitment pool, respondents completed a short pre-screen questionnaire which aimed to collect sociodemographic information (i.e., age, gender, and ethnicity) and contained questions pertaining to respondents' motivation for research, global motivation and personality. Their responses to the pre-screen questionnaire can be linked to their responses to subsequent studies using the respondents' anonymous code. After completing the pre-screen questionnaire, respondents could access the list of studies for which they were eligible. It is worth mentioning that none of the datasets used in the present research arise from studies with exclusion criteria. After signing up for any studies included in the present research, respondents were given access to an online survey to be completed at a moment of their choosing. Respondents received a 0.5% or 1% course credit, depending on the length of the survey, to be applied towards their final grade as a compensation for their participation in each study. To limit respondents' burden and selection bias, the present research relies on secondary use of data. The 23 datasets analyzed in the present research were obtained upon agreement with each principal investigator.

These datasets arise from five distinct research programs pertaining to academic resilience, the use of cellphone in the classroom, motivation and well-being, physical activity, and pro-environmental behaviours. The raw version of these datasets was provided, after data collection, by their respective principal investigator. Information relative to each research program, the number of datasets arising from each research program, the survey length, and the data collection period can be found in Appendix D.

The Multidimensional IER Detection Approach

In order to obtain a measure of IER we screened the data contained within each dataset using the multidimensional IER detection approach (mIERda) as per Gauthier and colleagues' (2025) methodology. Whereas the typical IER screening approaches require researchers to use multiple IER indicators and to establish a threshold beyond which one should be considered IER for each indicator, the mIERda relieves researchers from this responsibility by transferring the interpretation of the scores yielded by the IER indicators to an adaptation of an unsupervised learning algorithm (i.e., fair-cut forest). The mIERda was preferred over the more "classic" sequential approach as recent research has not only demonstrated the construct validity of the mIERda (Gauthier & Pelletier, 2025), but also found the mIERda to be more accurate than the sequential approach (Gauthier et al., 2025). This positions the mIERda as a performant, valid, and valuable IER screening approach.

Since the present study seeks to examine whether the determinants of IER differ at the start and the end of the survey, IER screening was performed using the scales containing the first 50 and last 50 items multiple-choice items of the survey. To measure IER at the start and the end of the survey, we computed the response coherence, response reliability, Mahalanobis distance, long string index, person-total correlation and intra-individual variability. Such approach was chosen since the length of the survey varies considerably from one study to another. Since the number of items included within the screening can impact its quality (Gauthier et al., 2025); using a similar number of items allowed us to ensure the resulting screening was comparable across datasets. Focusing on the first and last 50 items thus allows us to standardize the screening across datasets. We consequently propose to operationalize the length of the survey by using the number of items and by making the distinction between the first part of a survey and the last part

of the survey. This way, we can more accurately determine whether participants engage in IER throughout the entire survey or if they do it more at the end of longer surveys.

Since some indicators of IER depend on the scale's factorial structure, we did not split any scale and included as many scales as necessary to achieve a range of 47 to 54 items. After computing the indicators of IER, the data was treated by the fair-cut forest algorithm underpinning the mIERda which returned a score for each respondent. The optimal mIERda cutoff was then obtained and used to established respondents' IER status. Respondents with values smaller or equal to the cutoff were deemed attentive (0), whereas other respondents were deemed IER (1). The IER status was computed at the start and the end of the survey, for each respondent.

Substantive Measures

The main goal of the present study was to examine the determinants of IER across multiple datasets using a prospective longitudinal approach. To achieve this goal, we extracted the following personal and contextual determinants. More precisely, the personal determinants were measure prospectively (i.e., prior to participation in any study) whereas contextual determinants were collected during studies' completion and simultaneously to IER. For a visual representation of the data collection timeline, refer to Figure 4.1. The correlation between these determinants is reported at a within and between-level in Table 4.2.

Prospective Variables – Personal Determinants

Motivation Towards Research. Motivation for research participation was assessed using the Motivation Towards Research Scale (MTR; Cheung, 2019). The MTR is a 6-item measure which captures one's level of intrinsic ("Participating in research is interesting"), integrated ("I personally value the importance of being a research participant"), identified

(“Participating in research is important”), introjected (“I would feel guilty if I did not”), external (“This is a required component of my course”), and amotivation (“I do not know why I am participating in research; it is a waste of my time”) for research participation. Respondents were asked to which extent each statement corresponded to the reasons why they participate in research and responded to the items using a 7-point Likert scale ranging from 1 (*does not correspond at all*) to 7 (*corresponds entirely*). The intrinsic, integrated, and identified items were averaged to obtain a composite score of autonomous motivation ($M = 4.99$, $SD = 1.29$). The introjected and external items were averaged to obtain a score of controlled motivation ($M = 4.71$, $SD = 1.25$), while the amotivation item ($M = 1.75$, $SD = 1.13$) was interpreted as is. All resulting score ranged between 1 and 7.

Global Motivation. Respondents’ global motivation was measured using a 6-item version of the Global Motivation Scale (GMS; Cheung, 2019; Pelletier et al., 2004). Respondents are prompted to provide insight as of why they generally do things in their life (i.e., “In general, I do things...”). Then each item from the GMS captures one’s level of intrinsic (“for the pleasure of acquiring new knowledge”), integrated (“because they reflect what I value the most in life”), identified (“because I choose them as means to attain my objectives”), introjected (“because I would feel bad if I do not do them”), external (“in order to show others what I am capable of”), and amotivation (“although it does not make a difference whether I do them or not”), as defined by SDT. Respondents responded to the items using a 7-point Likert scale ranging from 1 (*not agree at all*) to 7 (*completely agree*). The intrinsic, integrated, and identified items were averaged to obtain a composite score of autonomous motivation ($M = 4.89$, $SD = 1.15$). The introjected and external items were averaged to obtain a score of controlled motivation

(*M* = 3.85, *SD* = 1.39), while the amotivation (*M* = 2.68, *SD* = 1.47) item was interpreted as is. All resulting score ranged between 1 and 7.

Personality. Respondents' personality was assessed using the 10-item version of the Big Five Inventory (BFI; Rammstedt & John, 2007). Respondents are prompted to report their perception (i.e., "I see myself as someone who ...") of their levels of agreeableness (e.g., "is generally trusting"; *M* = 3.59, *SD* = 0.86), conscientiousness (e.g., "does a thorough job"; *M* = 3.69, *SD* = 0.82), extraversion (e.g., "is reserved"; *M* = 3.08, *SD* = 1.04), openness to experience (e.g., "has an active imagination"; *M* = 3.53, *SD* = 0.86), and neuroticism (e.g., "gets nervous easily"; *M* = 3.34, *SD* = 1.10) using a Likert scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Each of these subscales were measured using two items. A composite score was obtained by reversing negative items (i.e., *items 1*, *3*, *4*, *5*, *7*) and by computing the average of each subscale's two items. All resulting score ranged between 1 and 5.

Contextual determinants

Study Topic. As it regards to the study topic, datasets were considered to share the same topic is they arose from the same research program and therefore pertained to the same main topic. The topic was established based on the theme of the broader research program from which the dataset originated. A detailed overview of the measure contained within each dataset can be found in Appendix D whereas descriptive information and statistics for each research program are summarized in Appendix E.

Length of the survey. For each dataset, the length of the survey was obtained by computing the number of items contained in the survey. The average survey length was of 189.37 (*SD* = 60.10) items and the length of the survey ranged between 108 and 299 items.

Time of the semester. In order to obtain a measure of when the respondent completed the study in any given semester, the time of the semester was measured by calculating the difference between the date of participation and the date at which the subject pool opened. At the university where the study took place, the subject pool opens 7 days after the start of classes. As such, the opening date can be computed by adding 7 days to the semester start date. Time of semester was treated as missing, and subsequently imputed, for datasets collected during the spring/summer semester as its length (5 weeks) substantially differs from the Fall/Winter semester (~13 weeks). In average, respondents participated 6.60 weeks after the pool opened at the start of the semester (SD = 3.59; range= 0.00 – 12.51). More information regarding the opening and closing date for each semester can be found in Appendix E.

Analytical Plan

Our analytical strategy followed various steps to explore the determinants of IER across numerous datasets.

Step 1 – IER screening. Before conducting any analysis, raw datasets were screened for IER. To do so, we used the methodology laid out by Gauthier and colleagues (2025) and performed the screening in *R* (version 4.4; R Core Team, 2024) using the *mIERda* package (Gauthier, 2025). For each dataset, as to obtain a measure of IER at the onset and end of the survey, using the aforementioned procedure. The resulting output was used as a measure of IER in the subsequent steps.

Step 2 – Extraction of the determinants. Once the IER screening was completed, meta-data was extracted from each dataset and used to generate an analytical dataset. More specifically, the length of the study, number of weeks since the beginning of the semester, and the study topic were extracted for each respondent and dataset. Afterwards, respondents' socio-

demographic information (i.e., age, gender, ethnicity), motivation towards research, global motivation, and personality were extracted from the pre-screen questionnaire and added to the analytical dataset. Missing data for motivation and personality measures were imputed with the *mice* package (version 3.18; Buuren & Groothuis-Oudshoorn, 2011) using single imputation.

Step 3 – Joint Bayesian multilevel logistic regression models. The examination of the personal determinants of IER carries unique challenges. Since IER is said to be an enduring interpersonal difference, it is likely that the responses to the personality, research, and global motivation measures provided by respondents flagged as IER in subsequent studies were also IER when they completed the pre-screen questionnaire. Nonetheless, performing IER screening on personality and motivation measures could introduce a collider bias in the analysis (Holmberg & Andersen, 2022). As such, reliance on a robust analytical approach was necessary to avoid needing to perform IER screening, while correcting the error introduced by these respondents. Relying on a multilevel Bayesian analytical approach allows for such model by recognizing that the baseline tendency to engage in IER differs across individuals and by yielding robust estimations of the effects of the predictors. Accordingly, a series of joint Bayesian multilevel logistic regression models were performed using a stepwise approach. In a first instance, a null model was estimated using the random effect of respondents. This was done to estimate the initial amount of within-person variable to be explained. In a second instance, motivation and contextual determinants (i.e., study topic, time of the semester, number of items) were added to the model. In a third instance, personality was added to the model to assess how it impacts the role played by motivation. Lastly, control variables (i.e., age group, gender, and ethnicity) were added to the model to assess how they impact the contribution of motivation and personality on IER.

The models were estimated using the brms package (Bürkner, 2018) in R, using Hamiltonian Monte Carlo (HMC) sampling with 4 Monte-Carlo Markov chains (MCMC), 5000 iterations per chain, and 2500 warmup iterations. Convergence of the MCMC was assessed using $\hat{R}$ (< 1.01) and effective sample size (ESS; > 1000). Posterior predictive checks were performed by examining visually the convergence of the MCMC and by computing the posterior predictive p-value (PPP) where a value closer to 0.50 is indicative that the model is representative of the data (Meng, 1994). After assessing the model convergence and adequacy, models were compared using cross-validation (ELPD LOO) and widely applicable information criterion (WAIC). Models with lower values were deemed to provide a better fit to the data (Vehtari et al., 2017).

As per Bayesian conventions, the posterior distributions were summarized by their medians and 89% credibility intervals (CrI). To facilitate interpretation, estimates were exponentiated to compute odds ratio. The meaningfulness of each predictor was assessed using the 89% highest density interval (HDI) and the ROPE (region of practical equivalence). More specifically, meaningfulness was assessed using the HDI+ROPE decision rule (Kruschke, 2018). Coefficients were deemed meaningful if the entirety of the 89% HDI was outside of the ROPE with a range representing +/- 20% in odds of IER in accordance with Cohen's (1988) definition of a negligible effect size (Kruschke, 2018). Similarly, we used the same criteria to assess whether the strength of each predictor varies at the start and the end of the survey.

Model Specification. Random intercepts were included for each respondent to capture individual variability in IER status. Given that IER status y at a moment t for a respondent i is a binary outcome, the sampling distribution is given by $y_{it}|p_{it} \sim \text{Bernoulli}(p_{it})$, where p_{it} is modeled as follows:

$$\text{logit}\left(\frac{p_{it}}{1-p_{it}}\right) = \beta_0 + \sum_j \beta_j X_{jit} + \mu_{0i}$$

where β_0 is the fixed intercept, β_j is the fixed regression coefficient for predictors X_j and $\mu_{0i} \sim$ *Normal* $(0, \sigma_u^2)$ denotes the random intercepts.

Prior Specifications and Sensitivity. To improve the stability of the model, while limiting priors' influence on the resulting estimation, we specified weakly informative priors as follows: Fixed regression coefficients (β_j) were assigned normal priors (*M* = 0, *SD* = 1). Such a prior was defined since it allows uncertainty about the direction and magnitude of each fixed effect while shrinking extreme estimates toward zero. Furthermore, such prior is useful in models containing numerous predictors as it limits overfitting and addresses instances of quasi-perfect or perfect separation (i.e., when a variable, or a set of variables, almost perfectly or perfectly classifies the participants in a logistic regression model; Simpson et al., 2017). While past research promotes the use of heavy-tailed priors (i.e., student's t, Cauchy distribution; Gelman et al., 2008), more recent research suggests using a normally distributed prior (*M* = 0, *SD* = 1) since heavy-tailed priors can artificially inflate regression coefficients in logistic regression (Simpson et al., 2017). The intercept parameters (β_0) were assigned a weakly informative normal prior (*M* = 0, *SD* = 10; Gelman et al., 2008). By imposing this prior, high credibility is granted to a probability of IER ranging between 0.01% and 99.0%. This prior thus informs the model that the prevalence of IER should fall within this range without influencing the baseline probability estimated by the model. The standard deviations (σ) of the random intercepts μ_{0i} were modelled with exponential (λ) priors, where λ = 1, to allow moderate between-person variability while avoiding unrealistically large variance (Simpson et al., 2017).

The likelihood and prior sensitivity of the posterior distribution were assessed using a power-scaling approach implemented using the *priorsense* package (Kallioinen et al., 2024). As

recommended by Kallioinen and colleagues (2014), priors were deemed suitable if the posterior distribution of each estimate was likelihood sensitive (> .05) and prior insensitive (< .05).

Results

Preliminary Analyses

Prior to performing the IER screening, each raw dataset was formatted and screened for out-of-range values. As IER screening relies on the use of raw data, no variables were winsorized nor recoded. Each dataset was screened for excessive missingness and respondents with over 30% missing data were excluded from the analysis. Afterwards, IER screening was performed using the mIERda.

The motivation and personality scales were then scored. Since respondents' age was collected using an ordinal scale which was limited to a range comprised between 16 years and younger and 30 years and older; responses were categorized as follows: 16-19 years old, 20-25 years old and 26 years and older to facilitate the analysis of the data. Afterwards, dummy variables were generated for categorical variables with more than two categories (i.e., gender, study topic, age group).

Then, variables that vary across time/studies were decomposed into within- and between-person components (i.e., time of the semester, survey length, study topic). This was done to disaggregate within- from between-person variance. The between-person component was computed by averaging each respondent's score for all observations. The within-person component was extracted by subtracting each respondent's mean score from their score at each observation. After disaggregating within and between effects, numerical variables were rescaled to a mean of 0 and standard deviation of 0.5. This was done, as per recommendation from

Gelman and colleagues (2008), to ensure that numerical and categorical determinants are measured and interpreted on a similar scale when using logistic regression.

Examination of the Determinants of IER

After assessing the prevalence of IER at a sample level, we conducted a series of joint Bayesian multilevel logistic regression. In a first instance, a null model was estimated. Then, at Step 1, motivation (i.e., amotivation, controlled, autonomous) at a global and research participation level was added as a predictor to the model. At Step 2, personality (i.e., openness to experience, conscientiousness, extraversion, agreeableness, neuroticism) was added to the model to assess its overlap with motivation. At Step 3, contextual variables (i.e., survey length, time of the semester, study topic) as well as covariates (i.e., gender, ethnicity, age group) were introduced to the model. At each step, we assessed whether the predictive strength of motivation was impacted by the addition of these variables. The results from each model are summarized in Table 4.3.

Model Adequacy

Estimation. Convergence diagnostics ($\hat{R}$ < 1.01, effective sample size > 1000/chains) indicated good mixing across all hypothesized models. Posterior predictive checks supported model adequacy (PPP ~ .50). Sensitivity analyses also revealed that the prior distribution was prior insensitive (<.05) but was likelihood sensitive (>.05) across all models, thus supporting that the prior selection was appropriate.

Null Model. The null model indicated an odds ratio of 0.40 with a respondent-level random intercept variance of 1.94 at the start of the survey. The null model indicated a baseline odds ratio of 0.50, with a respondent-level random intercept variance of 1.78 at the end of the survey. This highlights a notable inter-respondent variability in terms of their baseline odds of

IER at the start and the end of the survey. The intra-class correlation (ICC) indicated that 37.1% (start of the survey) and 35.1% (end of the survey) could be attributed to between-person differences, meaning most variability could be observed at a within-level (i.e., 1 – ICC). This justifies the use of a multilevel model.

Predictive Models. At Step 1, after adding research and global motivation variables, the model explained 10.8 % of the between-person variance observed at the start of the survey and 5.3% of the variance at the end of the survey. At Step 2, adding personality the model led to a 4% increase in explained between-person variance at the start and the end of the survey. After adding the contextual determinants (i.e., study topic, time of the semester, and survey length) and covariates (i.e., age, gender, ethnicity) the model (Step 3) explained an additional 7.4% in variance at the start of the survey and 5.9% at the end of the survey.

As variables were added to the model, the effect of motivation and personality variables did not meaningfully change (refer to Table 4.4), suggesting the effects were equivalent across steps. Consequently, solely the results for Step 3 are presented below. For a visual representation of the effect of personal and contextual determinants on IER at various points of the survey, refer to Figure 4.2 (start of the survey) and Figure 4.3 (end of the survey).

Effect of Personal Determinants

Effect of motivation for research participation. Results revealed that an increase in amotivation towards research participation equivalent to two standard deviations (*SD*) was associated with 55.7% higher odds of IER at the start of the survey and 46.9% higher odds of IER at the end of the survey. The effect was deemed meaningful since the entirety of the 89% HDI laid outside the ROPE (± 20% odds; OR range = [0.86, 1.19]). In contrast, the effect of controlled and autonomous motivation predictors for research participation on odds of IER was

uncertain at the start (Controlled: OR = 1.05, within ROPE = 96.4%; Autonomous: OR = 1.17, within ROPE = 60.8%) and the end (Controlled: OR = 1.11, within ROPE = 82.2%; autonomous OR = 1.14, within ROPE = 71.4%) of the survey. As such, the effects were deemed inconclusive.

Taken together, these results suggest that amotivation for research participation predicts greater odds of IER throughout the survey, supporting hypothesis 5a. However, the effects of autonomous and controlled motivation for research participation remain inconclusive, failing to support hypotheses 5b and 5c.

Effect of global motivation. Results were inconclusive as it regards the effect of global amotivation and controlled motivation throughout the survey. Although the analysis revealed that an increase equivalent to two standard deviations in global amotivation was associated with 30.5% higher odds of IER (within ROPE = 16.2%) at the start of the survey, and 11.9% higher odds of IER (within ROPE = 83.6%) at the end of the survey, the effects were found to be uncertain. This fails to provide support to hypothesis 6a.

A 2 SD increase in controlled motivation was associated with 34.0% higher odds of IER (within ROPE = 12.0%) at the start of the survey and 11.9% higher odds of IER (within ROPE = 78.0%) at the end of the survey. Since the 89% HDI does not fall completely within or outside the ROPE, the effect of global controlled motivation was hence deemed uncertain which fails to support hypothesis 6b.

While analyses failed to find any meaningful effect of global controlled motivation and global amotivation on odds of IER throughout the survey, results revealed a meaningful effect of global autonomous motivation at the start of the survey. More specifically, an increase of 2 SD in autonomous motivation led to 32.3% lower odds of IER (within ROPE = 0.0%) at the start of the survey. The meaningfulness of the effect however subsided at the end of the survey (OR = 0.85,

ROPE = 57.3%). These results offer partial support to hypothesis 6c as the effect was not constant throughout the survey. Taken together, these results suggest that global autonomous motivation seems to play a stronger role, in preventing IER, mainly at the start rather than at the end of the survey.

Hierarchical model of motivation. After extracting the effects of each form of motivation at a global and research participation level, we computed their differences in predictive strength. This was done to examine whether motivational variables were stronger predictors of IER at the domain-specific level (i.e., research participation) or at the dispositional level (i.e., global). As displayed in Table 4.5, the difference in relative strength was not meaningful between hierarchical levels of autonomous, controlled motivation, and amotivation. This suggests that the effect of each type of motivation on IER was not stronger at a domain-specific level compared to the effect observed at a global level which fails to support hypothesis 7.

Personality. As it regards to the effect of personality, the contribution varied by personality traits.

Openness to experience. Results revealed a negative and negligible effect of openness to experience in odds of engaging in IER at the start (OR = 0.86, within ROPE = 62.2%) and the end of the survey (OR = 0.91, within ROPE = 88.7%). Although uncertain, results suggest that greater levels of openness to experience might decrease engagement in IER. However, since uncertainty remains, results failed to provide evidence supporting hypothesis 4a.

Conscientiousness. Results suggest that respondents' levels of conscientiousness were a meaningful determinant of IER throughout the survey. More specifically, a 2 SD increase in conscientiousness was associated with 40% lower odds of engaging in IER at the start and 36.2%

lower odds of IER engagement at the end of the survey. The effects were deemed meaningful since the posterior distributions' 89% HDI laid completely outside the ROPE. These results suggest that greater levels of conscientiousness meaningfully decrease engagement in IER which provides support to hypothesis 4b.

Extraversion. Results were inconclusive regarding the effect of extraversion on the odds of engaging in IER at the start (OR = 0.94, within ROPE = 93.7%) and the end of the survey (OR = 0.77, within ROPE = 17.2%). While uncertainty remains, the results suggest that extraversion might decrease the odds of IER, but mostly at the end of the survey. Given the uncertainty remaining, this fails to support hypothesis 4c.

Agreeableness. Results reveal an inconclusive effect of agreeableness on engagement in IER at the start (OR = 0.82, within ROPE = 43.9%) and the end (OR = 0.87, within ROPE = 69.7%) of the survey. While the analysis suggest that agreeableness might decrease engagement in IER, the effect lacked meaningfulness which fails to support hypothesis 4d.

Neuroticism. Results were inconclusive relative to the effect of neuroticism on odds of IER at the start (OR = 1.10, within ROPE = 82.7%) and the end of the survey (OR = 1.18, within ROPE = 58.8%). While the analysis suggest that neuroticism might increase the odds of IER, the effect was uncertain. The results thus fail to support hypothesis 4e.

Effect of Contextual Determinants

Study topics. We examined if the odds of engaging in IER fluctuated, at a within-person level, across research programs (refer to Table 4.5 for detailed comparisons at a within- and between-person level). Results revealed no meaningful within-person differences in odds of IER across research programs (within ROPE = [17.1, 72.9%]). This suggests that the likelihood of a

respondent engaging in IER did not meaningfully change as they participated in studies on different topics.

At a between-person level, when examining whether the odds of engaging in IER at the start of the survey differed across study topics, the contrast analysis revealed two meaningful differences out of the 10 comparisons examined (within ROPE = [0.0, 40.9%]). This between-person effect of study topic was, however, limited at the start of the survey since the contrast analysis revealed no meaningful differences in odds of IER at the end of the survey (within ROPE = [11.9, 39.4%]). Overall, these results suggest that the effect of the study topics on odds of engaging in IER seems to differ depending on respondents' progress in the survey. The between-person effect observed at the start of the survey suggests that some topic might attract a profile of respondents who are more likely to engage in IER than others. Overall, results fail to support hypothesis 1a and partially support hypothesis 1b.

Time of the semester. At a within person-level, results revealed that the number of weeks since the start of the semester had no impact on engagement in IER at the start of the survey (OR = 1.04, within ROPE = 100%). The effect of the time of the semester was however inconclusive at the end of the survey (OR = 1.12, within ROPE = 79.3%). This suggests that, for a same respondent, participating in research earlier or later in the semester had no impact on their odds of IER at the start of the survey, and no meaningful impact on their odds of IER at the end of the survey. This failed to support hypothesis 3a.

At a between-person level, the effect of weeks since the start of the semester on odds of IER was inconclusive at the start (OR = 1.04, within ROPE = 97.7%) and the end (OR = 1.20, within ROPE = 48.5%) of the survey. Although results suggest that respondents who tend to participate in research later in the semester might have greater odds of engaging in IER at the

end of the survey, this effect was inconclusive. Overall, results suggest that respondents who tend to participate in research later in the semester are not meaningfully more likely to engage in IER throughout the survey than those who participate earlier. These results fail to support hypothesis 3b.

Survey length. At a within-person level, the effect of the length of the survey on odds of IER was inconclusive at the start (OR = 0.69, within ROPE = 21.6%) and the end (OR = 0.60, within ROPE = 6.02%) of the survey. This suggests that, for a same respondent, participating in studies containing more or less items did not meaningfully impact the likelihood that they would engage in IER. This failed to support hypothesis 2b.

At a between-person level, the results suggest a stronger negative effect of study length. Although the effect remained uncertain at the start (OR = 0.58, within ROPE = 6.3%), results unveiled a meaningful effect of study length at the end of the survey; where respondents who tend to participate in longer surveys were less likely to engage in IER (OR = 0.47, within ROPE = 0.0%). This suggest that respondents who tend to complete longer surveys have notably lower odds of engaging in IER at the end of the survey. This contradicts our hypothesis 2a.

Effect of Covariates

In order to ensure the robustness of the results, we controlled for the effect of gender, age, and ethnicity.

Gender. Results revealed that, compared to women, the odds of engaging in IER at the start of the survey were 27.5% greater for men (within ROPE = 7.9%) and 82.1% greater for individuals who identified otherwise (i.e., neither as a man nor a woman; within ROPE = 0.0%). These individuals also had 65.4% greater odds, compared to men, of engaging in IER at the start of the survey (within ROPE = 1.2%).

At the end of the survey, men and individuals who identified otherwise were both meaningfully more likely to engage in IER than women. More precisely, men had 39% greater odds of engaging in IER (within ROPE = 0.0%) compared to women. Individuals who identified otherwise displayed 54.1% greater odds of IER (within ROPE = 0.0%) compared to women. These individuals did not meaningfully differ from men in odds of IER at the end of the survey (OR = 1.05, within ROPE = 68.3%).

Overall, although uncertainty remains at the start of the survey, results suggest that women tend to have lower odds of IER than individuals identifying as men or otherwise, regardless of the moment of the survey. In contrast, men seemed less likely to engage in IER at the start of the survey than respondents who identified otherwise, but this effect subsided at the end of the survey.

Age. The effect of age on engagement in IER was uncertain throughout the survey. More precisely, respondents' age was not found to have a meaningful effect on the odds of IER at the start (ORs = [1.04, 1.41], within ROPE = [32.6, 36.9%]) and the end of the survey (ORs = [0.71, 0.96], within ROPE = [29.3, 97.1%]). This suggests that respondents' age did not meaningfully impact engagement in IER. For detailed contrasts, refer to Table 4.5.

Ethnicity. As it regards ethnicity, results suggest that respondents identifying as Black, Indigenous, and People of Colour (BIPOC) had meaningfully greater odds of engaging in IER throughout the survey. More precisely, the odds of engaging in IER at the start of the survey were 107% greater (within ROPE = 0.0%) for BIPOC respondents compared to White respondents. In contrast, the odds of engaging in IER were 86% greater at the end of the survey (within ROPE = 0.0%) for BIPOC respondents.

Discussion

Results from the present study shed light on the personal and contextual determinants of IER using a longitudinal prospective approach. Our objectives were not only to assess the role played by motivation (global & research participation) on IER, but also to assess its effect while accounting for the effect of other known determinants of IER. Additionally, we sought to examine whether the effect of these determinants differed as participants progressed into the survey (i.e., at the start vs. the end of the survey).

Effect of Motivation

In a first instance, the present study sheds light on the central role played by respondents' motivation at a domain-specific (i.e., research participation) and dispositional level (i.e., global) on their subsequent odds of engaging in IER. As predicted, results indicated that as respondents exhibited greater levels of amotivation towards research participation, they were also more likely to engage in IER not only at the start, but also at the end of the survey. More specifically, this suggests that respondents who approach research participation in an amotivated manner (i.e., without seeing or believing in its purpose or value) are more likely to engage in IER regardless of the moment of the survey. Interestingly, at a global level, amotivation was not found to meaningfully predict IER at the start and the end of the survey. This implies that respondents who have a higher global tendency of being amotivated are not more likely to engage in IER, which was contrary to our hypothesis and results from Gauthier & Pelletier (2025) who reported a positive correlation between global amotivation and engagement in IER throughout the survey.

Even if results from the present study failed to corroborate our hypothesis, the absence of meaningful effect of global amotivation on IER remains coherent with results arising from other research which demonstrated that although global amotivation plays a central role in

understanding respondents' engagement in IER, more specific forms of amotivation (e.g., towards research participation) serves as a more useful predictor of IER (Cheung, 2019). While our findings did not reveal any credible difference between the effect of global and research amotivation in terms of odds of IER, it remains that solely research amotivation meaningfully predicted more IER. Although we did not test the effect of global amotivation on research amotivation, as global amotivation denotes of a general deposition, it is possible that global amotivation indirectly impacts engagement in IER via research amotivation which not only could explain its non-meaningful direct effect but would also be in line with results reported by Cheung (2019). It is also possible that the non-meaningfulness of the effect of global amotivation on IER might be due to the nature of the concept at hand. In other words, since global amotivation refers to a global lack of intention to act, respondents with higher global tendency of amotivation may simply be less likely to engage in research at all, which would prevent them from engaging in IER.

Additionally, our results further support the evidence arising from Cheung (2019) and Gauthier & Pelletier (2025) stating that controlled motivation (at a global and research participation level) was not a meaningful predictor of IER. Results also suggest that higher levels of autonomous motivation at a global level, but not at a research participation level, predicted decreased odds of IER at the start of the survey. Interestingly, this effect lost its meaningfulness at the end of the survey; in such instance global autonomous motivation did not meaningfully predict IER. Although consistent with past research indicating a negative association between global autonomous motivation and IER (Cheung, 2019), results from the present study brings nuance by demonstrating that the meaningfulness of its effect fluctuates across the survey. Given that global motivation is a measure of one's general motivational orientations and tendencies

(Vallerand, 1997), global autonomous motivation might influence respondents' disposition and how they approach the task at hand more strongly than their engagement during the task. This suggests that participants with stronger global autonomous orientations may be more likely to be attentive, at least at the start of the survey, across studies. Furthermore, such finding is also coherent with recent research indicating that autonomous motivation tends to be more specific and unstable than controlled forms of motivation (Chanal et al., 2025). Consequently, while global autonomous motivation might decrease IER at the onset, it is possible that its relative importance and contribution decrease as the specificity of the task increases in salience. Similarly, this might also explain why the effect of research amotivation on respondents' engagement in IER remains stable from the onset to the end of the survey.

Personality

The present study sheds light on a mixed effect of personality on IER. Contrarily to what has been hypothesized, amongst all the personality traits (i.e., agreeableness, conscientiousness, extraversion, openness to experience, neuroticism) considered in the present study, solely conscientiousness was a meaningful determinant of IER. Results failed to provide credible evidence of the predictive effect of respondents' openness to experience, extraversion, agreeableness, and neuroticism on IER. Precisely, solely respondents' level of conscientiousness was associated with decreased odds of IER at the start and the end of the survey. Such finding partially aligns with results from previous studies not only reporting a negative association between conscientiousness and IER, but also a negative association between IER and agreeableness, extraversion, and emotional stability (Bowling et al., 2016; Dunn et al., 2018; Grau et al., 2019; Ward et al., 2017). On a larger scale, this divergence from the literature means that IER might be less tied to personality as initially believed which corroborates results from

other studies reporting a null association between IER and personality (Camus, 2015; Harms et al., 2017; Hasselhorn et al., 2025). Such result nonetheless aligns with findings from past research suggesting that amongst personality traits, conscientiousness is the strongest predictor of IER (Mazza & Arthur, 2025).

Taken together, such findings have serious implications. Given that IER respondents often get removed from samples, a strong association between personality and IER would have suggested that IER screening removes a large body of respondents with specific personality traits. Doing so inherently implies decreasing the diversity of research samples. The present findings, however, provide a nuanced perspective on this matter. While conscientiousness remained a stable and meaningful determinant of IER, this effect might not be immutable. Research arising from the field of SDT suggests that since individuals with greater levels of conscientiousness tend to be more self-aware and less impulsive, they might also be more likely to take decisions and engage in activities that align with their needs and wants (Ryan et al., 2019). It is this very awareness and congruence to the self that is necessary to the development of a sense of self-determination. Similarly, conscientiousness is associated with less need frustration and more need satisfaction (Story et al., 2025) and is linked to higher levels of autonomous motivation (Tu et al., 2020). In the context of a university-based research recruitment pool, the context in which research participation takes place might be perceived as controlling, thus need frustrating. Since individuals with low conscientiousness may be more impulsive, intolerant to delayed gratification and prone to engage in defiance in the face of need frustration (Ryan et al., 2019), it is likely that respondents' needs satisfaction moderates the association between conscientiousness and IER. Furthermore, a study revealed that individuals tend to be more open, conscientious, agreeable, extroverted and emotionally stable when

interacting with people who are supportive of their basic psychological needs (Lynch et al., 2009); suggesting potentially relevant intervention targets.

Time of the Semester

The present results bring nuance to the literature by demonstrating that respondents who, in average, completed their research participation component later in the semester might be more likely to engage in IER at the end of the survey. Similarly, those who completed a study later than usual in the semester might be more likely to engage IER at the end of the survey. Results should, however, be interpreted cautiously. While the effects of the time of the semester (at a within- and between-person level) were in the expected direction; they were not sufficiently large to be deemed meaningful neither at the start nor the end of the survey. Nonetheless, these results are interesting when we consider that the effect of the time of the semester was almost null at the start of the survey and became stronger at the end of the survey. This means that fatigue could play a role which corroborates results from studies demonstrating the detrimental effect of respondent fatigue on data quality (Jeong et al., 2023). These results align with research reporting a significant yet small effect of the time of the semester on the prevalence of IER (Cheung, 2019; Nicholls et al., 2015). It is also plausible that respondents' motivation interacts with time of the semester since previous studies have demonstrated that participants that are more engaged (Witt et al., 2011) and have greater levels of intrinsic motivation (Nicholls et al., 2015) tend to participate in research earlier in the semester. Overall, results suggest that even if the time of the semester does influence the odds of IER, some determinants might be more important or meaningful to consider.

Survey Length

Contrarily to what had been hypothesized, survey length was negatively associated with odds of IER. Interestingly, respondents who tended to complete longer surveys than average were less likely than others to engage in IER at the end of the survey. This effect was, however, not meaningful at the start of the survey. Similarly, while the effect was not deemed meaningful, respondents tended had lower odds of IER when they completed longer studies, irrespective of the study topic. This fails to support past literature stating that IER is more prevalent in longer studies (Bowling et al., 2021; Brower, 2018). Although these results should be interpreted with caution, as a lack of experimental design prevents us from establishing formal causality, it is possible that respondents deciding to partake in longer studies might do so more intentionally, hence limiting their odds of IER. Simultaneously, it is also possible that respondents who are already disengaged might try to limit their engagement to the minimum and be more likely to sign up and participate in shorter studies. There is also a possibility that participants who experience lower levels of amotivation might also be more likely to sign up for longer studies which could translate in less disengagement (i.e., IER). This could also explain why within-person fluctuations are also observed, it is likely that respondents might try to modulate their engagement across time as their dispositional resources fluctuate.

Survey Topic

While past research has demonstrated that lower interest in study topics is associated with increased odds of IER in longer surveys (Brower, 2018; Ward & Meade, 2023), no study had investigated whether the prevalence fluctuated in studies of various topics. Results revealed that participants' odds of engaging in IER did not meaningfully differ as they participated in studies of different topics. While there was no meaningful effect of study topic at a within-person level,

multiple comparisons of the between-person effect of research topics reveal two meaningful differences, at the start of the survey, out of the 10 examined comparisons. These differences were, however, not observed at the end of the survey. This implies that although some study topics can be more attractive to participants who might be more prone to engaging in IER at the start of the survey, the topic nonetheless seemed to have a limited effect. Overall, the lack of meaningful within-person differences in odds of IER across study topics suggests that IER mostly results from inter-individual differences. Future research would benefit from examining whether interest, or even motivation, mediates the association between the study topic and IER. Since the present research only considered datasets arising from 5 distinct research programs, future research would also benefit from investigating differences across a broader amount of study topics.

Covariates

To ensure the generalizability of our results, we controlled for multiple covariates. As it regards to the effect of gender, women were less likely to engage in IER than men (i.e., at the end of the survey) or respondents identifying otherwise (i.e., throughout the survey). Although the effect was uncertain, men had lower odds of IER than respondents identifying otherwise at the start of the survey, but this difference subsided at the end of the survey. These results are in line with past research reporting that men are more likely to engage in IER than women (Braitman et al., 2025; Cheung, 2019; Kuang et al., 2025; Schneider et al., 2018). As it pertains to the effect of age, results fail to reveal any meaningful difference in odds of IER between respondents aged 16-19, 20-25, and those 26 years and older, which supports past literature reporting null effects of age on IER (Grau et al., 2019). Lastly, results suggest that respondents who identified as BIPOC were meaningfully more likely to engage in IER throughout the survey

than respondents who identified as White. This finding aligns with recent studies which have reported cross-cultural differences in terms of prevalence of IER (Braitman et al., 2025; Grau et al., 2019) and supports the relevance of accounting for respondents' ethnicity when examining the determinants of IER.

Limitations and Future Research

Despite the insight provided by the present study, some limitations remain. For example, we did not rely on an experimental design to examine the effect of contextual determinants. While relying on a prospective longitudinal approach allowed us to distinguish within- and between-person effects, future research would benefit from implementing experimental manipulations (e.g., random attribution to a topic, or random attribution to a short vs. long survey condition) rather than letting participants sign up for the study of their choice. Furthermore, future research might benefit from examining a broader number of study topics. While we failed to find any meaningful difference in odds of IER across topics, we relied on a limited number of topics arising from the field of academic motivation, education, physical activity, and environmental psychology. As such, the surveys did not cover the breath of the field of psychology, never mind other disciplines relying on surveys. Future research would benefit from examining whether these findings extend to other disciplines and sub-disciplines. Last but not least, the present research relied on self-report measures of global, research participation motivation, and personality to predict the occurrence of IER. Given the known detrimental impact of IER on data quality and the conclusions that can be drawn from survey data, the present conclusions might be biased by the presence of IER in our sample. While we relied (a) on a Bayesian analytical approach to regularize the results and limit the risk of bias and (b) on a random intercept approach to model and account for respondent-related variance (i.e., error),

future research would benefit from relying on a multi-rater approach, laboratory observations, or additional observations.

Conclusion

Overall, the present research sheds light on the associations between IER and its contextual and personal determinants across time. This study suggests that IER shared a limited association with contextual determinants and was more strongly related to personal determinants throughout the survey. Nonetheless, more research is needed to understand the factors underpinning IER. Taken together, the present study provides support of the validity of the mIERda, highlights the relevance of considering the role played by motivation at various hierarchical levels, and positions Self-Determination Theory as a useful framework to understand why respondents engage in IER.

References

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*(6), 2489–2505. https://doi.org/10.3758/s13428-020-01401-8

Ashley, M., & Shaughnessy, K. (2021). Predicting Insufficient Effort Responding: The Relation Between Negative Thoughts, Emotions, and Online Survey Responses. *Canadian Journal of Behavioural Science*, *55*(3). 1–13. https://doi.org/10.1037/cbs0000308

Aviv, A. L., Zelenski, J. M., Rallo, L., & Larsen, R. J. (2002). Who comes when: personality differences in early and later participation in a university subject pool. *Personality and Individual Differences*, *33*(3), 487–496. https://doi.org/10.1016/S0191-8869(01)00199-4

Baer, R. A., Ballenger, J., Berry, D. T., & Wetter, M. W. (1997). Detection of Random Responding on the MMPI-A. *Journal of Personality Assessment*, *68*(1), 139–151. https://doi.org/10.1207/s15327752jpa6801_11

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*, *4*(3), 340–345. https://doi.org/10.1037/1040-3590.4.3.340

Bowling, N. A., Gibson, A. M., Houpt, J. W., & Brower, C. K. (2021). Will the Questions Ever End? Person-Level Increases in Careless Responding During Questionnaire Completion. *Organizational Research Methods*, *24*(4), 718–738. https://doi.org/10.1177/1094428120947794

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who Cares and Who Is Careless? Insufficient Effort Responding as a Reflection of

Respondent Personality. *Journal of Personality and Social Psychology*, *111*(2), 218–229. http://dx.doi.org/10.1037/pspp0000085.supp

Braitman, A. L., Petrey, A. M., Shipley, J. L., Ayala Guzman, R., Renzoni, E., Looby, A., & Bravo, A. J. (2025). Check your data before you wreck your model: The impact of careless responding on substance use data quality. *Alcohol, Clinical and Experimental Research*, *49*(4), 941–951. https://doi.org/10.1111/acer.70024

Brower, C. K. (2018). *Too Long and Too Boring: The Effects of Survey Length and Too Long and Too Boring: The Effects of Survey Length and Interest on Careless Responding Interest* [Wright State University, Master's Thesis]. Core Scholar. https://corescholar.libraries.wright.edu/etd_all/1918

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3). https://doi.org/10.18637/jss.v045.i03

Camus, K. A. (2015). *Once Careless, Always Careless? Temporal and Situational Once Careless, Always Careless? Temporal and Situational Stability of Insufficient Effort Responding (IER) Stability of Insufficient Effort Responding (IER)* [Masters Thesis, Wright State University]. https://corescholar.libraries.wright.edu/etd_all

Chanal, J., Remolif, B., & Paumier, D. (2025). On the hierarchical and situational multidimensionality of motivation according to self-determination theory: further investigating the specificity hypothesis. *Motivation and Emotion*. https://doi.org/10.1007/s11031-025-10139-9

Cheung, S. (2019). *Understanding Insufficient Effort Responding from a Self-Determination Theory Perspective* [Doctoral Thesis, University of Ottawa]. uO Research. https://ruor.uottawa.ca/items/5db7e905-2c10-42a6-a335-6bf73341c438

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edition). Lawrence Erlbaum Associates

de Calvo, M. P. C., & Reich, D. A. (2007). Spontaneous correction in the behavioral confirmation process: The role of naturally-occurring variations in self-regulatory resources. *Basic and Applied Social Psychology*, *29*, 351–364. https://doi.org/10.1080/01973530701665132

Deci, E., & Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer New York. https://doi.org/10.1007/978-1-4899-2271-7

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences. *Journal of Business and Psychology*, *33*(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Francavilla, N. M., Meade, A. W., & Young, A. L. (2019). Social Interaction and Internet-Based Surveys: Examining the Effects of Virtual and In-Person Proctors on Careless Response. *Applied Psychology*, *68*(2), 223–249. https://doi.org/10.1111/apps.12159

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Gauthier, A. J. (2025). *The multidimensional Insufficient Effort Responding detection approach (mIERda)*. https://github.com/AJGauthier/mIERda.

Gauthier, A. J., & Pelletier, L. G. (2025). *A Multi-Study Examination of the Nomological Validity of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)*[Unpublished manuscript]. School of Psychology, University of Ottawa.

Gauthier, A. J., Pelletier, L. G., & Gravel, C. (2025). *Sorting the Good from the mIERda – Introducing and Testing the performance of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)*[Unpublished manuscript]. School of Psychology, University of Ottawa.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4). https://doi.org/10.1214/08-AOAS191

Gibson, A. M., & Bowling, N. A. (2020). The Effects of Questionnaire Length and Behavioral Consequences on Careless Responding. *European Journal of Psychological Assessment*, *36*(2), 410–420. https://doi.org/10.1027/1015-5759/a000526

Grau, I., Ebbeler, C., & Banse, R. (2019). Cultural Differences in Careless Responding. *Journal of Cross-Cultural Psychology*, *50*(3), 336–357. https://doi.org/10.1177/0022022119827379

Green-Demers, I., Pelletier, L. G., & Ménard, S. (1997). The impact of behavioural difficulty on the saliency of the association between self-determined motivation and environmental behaviours. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 29*(3), 157–166. https://doi.org/10.1037/0008-400X.29.3.157

Guertin, C., Barbeau, K., Pelletier, L., & Martinelli, G. (2017). Why do women engage in fat talk? Examining fat talk using Self-Determination Theory as an explanatory framework. *Body Image*, *20*, 7–15. https://doi.org/10.1016/j.bodyim.2016.10.008

Harms, C., Jackel, L., & Montag, C. (2017). Reliability and completion speed in online questionnaires under consideration of personality. *Personality and Individual Differences*, *111*, 281–290. https://doi.org/10.1016/j.paid.2017.02.015

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2025). Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: Factors influencing careless responding. *Psychological Methods*, *30*(2), 374–392. https://doi.org/10.1037/met0000580

Herzog, A. R., & Bachman, J. G. (1981). Effects of Questionnaire Length on Response Quality. *The Public Opinion Quarterly*, *45*(4), 549–559. https://doi.org/10.1086/268687

Holmberg, M. J., & Andersen, L. W. (2022). Collider Bias. *JAMA*, *327*(13), 1282. https://doi.org/10.1001/jama.2022.1820

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and Deterring Insufficient Effort Responding to Surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., & DeSimone, J. A. (2021). Insufficient Effort Responding as a Potential Confound between Survey Measures and Objective Tests. *Journal of Business and Psychology*, *36*, 807–828. https://doi.org/10.1007/s10869-020-09707-2

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

Jeong, D., Aggarwal, S., Robinson, J., Kumar, N., Spearot, A., & Park, D. S. (2023). Exhaustive or exhausting? Evidence on respondent fatigue in long surveys. *Journal of Development Economics*, *161*, 102992. https://doi.org/10.1016/j.jdeveco.2022.102992

Kallioinen, N., Paananen, T., Bürkner, P.-C., & Vehtari, A. (2024). Detecting and diagnosing prior and likelihood sensitivity with power-scaling. *Statistics and Computing*, *34*(1), 57. https://doi.org/10.1007/s11222-023-10366-5

Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, *10*(JUN). https://doi.org/10.3389/fpsyg.2019.01258

Kam, C. C. S., & Meyer, J. P. (2015). How Careless Responding and Acquiescence Response Bias Can Influence Construct Dimensionality: The Case of Job Satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Kuang, H., Zhu, L., Yin, H., Zhang, Z., Jing, B., & Kuang, J. (2025). The Impact of Individual Factors on Careless Responding Across Different Mental Disorder Screenings: Cross-Sectional Study. *Journal of Medical Internet Research*, *27*, e70451–e70451. https://doi.org/10.2196/70451

Lavigne, G. L., & Vallerand, R. J. (2010). The Dynamic Processes of Influence Between Contextual and Situational Motivation: A Test of the Hierarchical Model in a Science Education Setting. *Journal of Applied Social Psychology*, *40*(9), 2343–2359. https://doi.org/10.1111/j.1559-1816.2010.00661

Lynch, M. F., La Guardia, J. G., & Ryan, R. M. (2009). On being yourself in different cultures: ideal and actual self-concept, autonomy support, and well-being in China, Russia, and the United States. *The Journal of Positive Psychology*, *4*(4), 290–304. https://doi.org/10.1080/17439760902933765

Mazza, N., & Arthur, W., Jr. (2025). Correlates of Careless Responding: Trait and State Antecedents and Criteria. *Human Performance*, *38*(5), 243–264. https://doi.org/10.1080/08959285.2025.2552213

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*(1), 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Meng, X.-L. (1994). Posterior Predictive p-Values. *The Annals of Statistics*, *22*(3). https://doi.org/10.1214/aos/1176325622

Nicholls, M. E. R., Loveless, K. M., Thomas, N. A., Loetscher, T., & Churches, O. (2015). Some participants may be better than others: Sustained attention and motivation are higher early in semester. *Quarterly Journal of Experimental Psychology*, *68*(1), 10–18. https://doi.org/10.1080/17470218.2014.925481

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom. *Theory and Research in Education*, *7*(2), 133–144. https://doi.org/10.1177/1477878509104318

Ntoumanis, N., Ng, J. Y. Y., Prestwich, A., Quested, E., Hancox, J. E., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Lonsdale, C., & Williams, G. C. (2021). A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. *Health Psychology Review*, *15*(2), 214–244. https://doi.org/10.1080/17437199.2020.1718529

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Pelletier, L. G., Dion, S. C., Slovinec-D'Angelo, M., & Reid, R. (2004). Why do you regulate what you eat? Relationships between forms of regulation, eating behaviors, sustained dietary behavior change, and psychological adjustment. *Motivation and Emotion*, *28*(3), 245–277. https://doi.org/10.1023/B:MOEM.0000040154.40922.14

Pelletier, L. G., Tuson, K. M., Green-Demers, I., Noels, K., & Beaton, A. M. (1998). Why are you doing things for the environment? The Motivation Toward the Environment Scale (MTES). *Journal of Applied Social Psychology*, *28*(5), 437–468. https://doi.org/10.1111/j.1559-1816.1998.tb01714.x

Pelletier, L. G., & Rocchi, M. (2023). Organismic integration theory: A theory of regulatory styles, internalization, integration, and human functioning in society. In R. M. Ryan (Ed.), *The Oxford handbook of self-determination theory* (pp. 53–83). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197600047.013.4

R Core Team. (2024). *R: A language and environment for statistical computing* (4.4.0). R Foundation for Statistical Computing.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10 item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Ryan, R. M. (Ed.). (2023). *The Oxford handbook of self-determination theory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197600047.001.0001

Ryan, R. M., & Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. The Guilford Press. https://doi.org/10.1521/978.14625/28806

Ryan, R. M., Soenens, B., & Vansteenkiste, M. (2019). Reflections on self-determination theory as an organizing framework for personality psychology: Interfaces, integrations, issues, and unfinished business. *Journal of Personality*, *87*(1), 115–145. https://doi.org/10.1111/jopy.12440

Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research*, *27*(4), 1077–1088. https://doi.org/10.1007/s11136-017-1767-2

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, *32*(1). https://doi.org/10.1214/16-STS576

Spence, G. B., & Oades, L. G. (2011). Coaching with self-determination in mind: Using theory to advance evidence-based coaching practice. *International Journal of Evidence Based Coaching and Mentoring*, *9*(2), 37–55.

Story, P. A., Bailey, Z., & Foo, N. G. (2025). The role of conscientiousness in meeting psychological needs: A whole trait theory perspective. *Personality and Individual Differences*, *240*, 113138. https://doi.org/10.1016/j.paid.2025.113138

Tu, Y., Lu, X., Wang, S., & Liu, Y. (2020). When and why conscientious employees are proactive: A three-wave investigation on employees' conscientiousness and organizational proactive behavior. *Personality and Individual Differences*, *159*, 109865. https://doi.org/10.1016/j.paid.2020.109865

Vallerand, R. J. (1997). Toward a Hierarchical Model of Intrinsic and Extrinsic Motivation. *Advances in Experimental Social Psychology*, *29*, 271–360. https://doi.org/10.1016/S0065-2601(08)60019-2

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417–430. https://doi.org/10.1016/J.CHB.2017.06.032

Ward, MK., & Meade, A. W. (2023). Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices. *Annual Review of Psychology*, *74*(1). https://doi.org/10.1146/annurev-psych-040422-045007

Ward, MK., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417–430. https://doi.org/10.1016/J.CHB.2017.06.032

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender Differences in Personality across the Ten Aspects of the Big Five. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00178

Williams, G. C., Grow, V. M., Freedman, Z. R., Ryan, R. M., & Deci, E. L. (1996). Motivational predictors of weight loss and weight-loss maintenance. *Journal of Personality and Social Psychology*, *70*(1), 115–126. https://doi.org/10.1037/0022-3514.70.1.115

Witt, E. A., Donnellan, M. B., & Orlando, M. J. (2011). Timing and selection effects within a psychology subject pool: Personality and sex matter. *Personality and Individual Differences*, *50*(3), 355–359. https://doi.org/10.1016/J.PAID.2010.10.019

Woods, C. M. (2006). Careless Responding to Reverse-Worded Items: Implications for Confirmatory Factor Analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3). https://doi.org/10.1007/s10862-005-9004-7

Table 4.1

Aggregated Socio-Demographic Information

Variables		*n*	%
Status – Start of survey*			
	Attentive	3351	66.6
	IER	1679	33.4
Status – End of survey*			
	Attentive	3152	62.7
	IER	1878	37.3
Gender			
	Women	1501	65.8
	Men	517	22.7
	Other	163	7.1
	Undisclosed	101	4.4
Age*			
	16-19 years	1558	67.9
	20-25 years	667	29.1
	26 years and older	36	1.6
	Undisclosed	35	1.4
Ethnicity			
	BIPOC	1198	54.1
	White	956	43.1
	Undisclosed	62	2.8

Note. The values reported for *n* and % denotes the number and proportion of distinct observations, not the number of distinct respondents. The proportion might not sum to 100% as some observations are missing. *The sum for these variables does not sum to $N = 2216$ since participants' status and age changed across time. BIPOC = Black, Indigenous and People of Colour

Table 4.2

Correlation Matrix Between Variables of Interest

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Within-Person (Level 1)																				
1 – IER (Start Survey)	–																			
2 – IER (End Survey)	.45*	–																		
3 – Survey Length	-.06*	-.06*	–																	
4 – Time of the Semester	.01	.02	-.11*	–																
5 – Research Program 1	-.05*	-.03	.58*	.02	–															
6 – Research Program 2	-.01	.01	-.40*	.07*	-.04*	–														
7 – Research Program 3	.01	-.01	.34*	-.06*	-.07*	-.81*	–													
8 – Research Program 4	.07*	.06*	-.79*	.11*	-.51*	-.03	-.06*	–												
9 – Research Program 5	-.05*	-.05*	.54*	-.15*	-.07*	-.02	-.04*	-.75*	–											
Between-Person (Level 2)																				
1 – Mean IER (Start Survey)	–																			
2 – Mean IER (End Survey)	.41*	–																		
3 – Survey Length	-.11*	-.13*	–																	
4 – Time of the Semester	.04	.08*	.05*	–																
5 – Research Program 1	-.11*	-.11*	.95*	-.05*	–															
6 – Research Program 2	-.01	-.03	-.12*	.12*	-.28*	–														
7 – Research Program 3	.01	-.03	-.09*	.09*	-.30*	.71*	–													
8 – Research Program 4	.07*	.13*	-.65*	.06*	-.48*	-.47*	-.48*	–												
9 – Research Program 5	.08*	.05*	-.53	-.17*	-.51*	-.31*	-.32*	.26*	–											
10 – Amotivation RM	.17*	.15*	-.04*	.11*	-.06*	.00	.00	.03	.05*	–										
11 – Controlled RM	.02	.02	-.02	-.02	-.02	.01	.00	.01	.02	.03	–									
12 – Autonomous RM	-.07*	-.05*	.06*	-.07*	.07*	-.01	.00	-.04*	-.05*	-.37*	.18*	–								
13 – Amotivation GM	.13*	.09*	-.05*	.04*	-.05*	-.02	-.01	.04*	.06*	.24*	.07*	-.09*	–							
14 – Controlled GM	.06*	.04	.01	.00	.00	-.03	-.02	.00	.04*	.05*	.31*	.25*	.16*	–						
15 – Autonomous GM	-.12*	-.08*	.03	-.03	.02	.03	.01	-.04*	-.02	-.16*	.14*	.42*	-.09*	.33*	–					
16 – Openness	-.04	-.03	-.03	.01	-.03	-.01	.01	.04*	.00	-.05*	-.02	.06*	-.07*	-.02	.07*	–				
17 – Conscientiousness	-.19*	-.17*	.02	-.12*	.02	.01	.01	-.07*	.03	-.15*	.06*	.20*	-.11*	.00	.30*	.03	–			
18 – Extraversion	.02	-.03	.00	-.06*	.01	-.02	-.01	.00	.00	.02	-.04	-.10*	.03	-.06*	-.09*	-.06*	-.13*	–		
19 – Agreeableness	-.07*	-.06*	.05*	-.02	.05*	-.01	-.03	-.02	-.02	-.10*	.01	.15*	-.04*	.00	.10*	.00	.19*	-.16*	–	
20 – Neuroticism	.01	.01	.03	-.03	.03	-.01	.01	-.04*	.00	-.04	.04*	.02	.02	.15*	-.08*	.01	-.20*	.22*	-.14*	–

Note. $N_{Within\text{-}Person}$ = 5030, $N_{Between\text{-}Person}$ = 2216. All variables were rescaled to $M = 0$, $SD = 0.5$. *$p < .01$. Research program 1 = Academic Resilience, Research program 2 = Cellphone use in Classroom, Research program 3 = Motivation and Well-being, Research program 4 = Physical Activity, Research program 5 = Pro-environmental behaviours.

Table 4.3

Stepwise Multilevel Logistic Regression Results

Determinants	Null Model		Step 1		Step 2		Step 3	
	Start of Survey	End of Survey	Start of Survey	End of Survey	Start of Survey	End of Survey	Start of survey	End of survey
Intercept	0.40 [0.37 – 0.43]	0.50 [0.47 – 0.54]	0.40 [0.37 – 0.43]	0.50 [0.47 – 0.54]	0.40 [0.36 – 0.43]	0.50 [0.47 – 0.54]	0.35 [0.18 – 0.53]	0.59 [0.31 – 0.92]
				Within-Level				
Contextual Determinants								
Survey length	–		–	–	–	–	0.69 [0.41 – 1.01]	0.60 [0.37 – 0.87]
Time of the semester	–		–	–	–	–	*0.98 [0.83, 1.15]*	1.12 [0.95, 1.30]
Study Topic								
Research Program 1 vs.								
Research Program 2	–		–	–	–	–	1.03 [0.42 – 1.79]	0.82 [0.35, 1.38]
Research Program 3	–		–	–	–	–	1.53 [0.80 – 2.38]	0.96 [0.51, 1.45]
Research Program 4	–		–	–	–	–	1.32 [0.79 – 1.95]	1.05 [0.65, 1.52]
Research Program 5	–		–	–	–	–	1.01 [0.71 – 1.34]	0.97 [0.70, 1.27]
				Between-Level				
Contextual Determinants								
Survey length	–		–	–	–	–	0.57 [0.31, 0.87]	**0.47 [0.26, 0.72]**
Time of the semester	–		–	–	–	–	1.04 [0.87 – 1.20]	1.20 [1.02, 1.40]
Study Topic								
Research Program 1 vs.	–		–	–	–	–		
Research Program 2	–		–	–	–	–	0.42 [0.13, 0.81]	0.51 [0.16, 0.97]

Determinants	Null Model		Step 1		Step 2		Step 3	
	Start of Survey	End of Survey	Start of Survey	End of Survey	Start of Survey	End of Survey	Start of survey	End of survey
Research Program 3	–		–	–	–	–	1.16 [0.48, 2.04]	0.77 [0.29, 1.33]
Research Program 4	–		–	–	–	–	0.55 [0.18, 1.03]	0.90 [0.32, 1.67]
Research Program 5	–		–	–	–	–	1.26 [0.53, 2.12]	0.59 [0.27, 1.00]
Personal Determinants								
Global Motivation								
Autonomous Motivation	–		**0.57 [0.48, 0.68]**	**0.69 [0.59, 0.81]**	**0.67 [0.57, 0.80]**	0.81 [0.69, 0.96]	**0.68 [0.55, 0.80]**	0.85 [0.71, 1.00]
Controlled Motivation	–		1.37 [1.17, 1.62]	1.18 [1.01, 1.39]	1.33 [1.13, 1.57]	1.10 [0.94, 1.31]	1.34 [1.09, 1.57]	1.12 [0.93, 1.30]
Amotivation	–		1.38 [1.19, 1.61]	1.20 [1.03, 1.39]	1.33 [1.14, 1.55]	1.16 [1.01, 1.35]	1.31 [1.10, 1.51]	1.11 [0.95, 1.28]
Research Motivation								
Autonomous Motivation	–		1.10 [0.92, 1.31]	1.11 [0.94, 1.31]	1.17 [10,98, 1.39]	1.15 [0.99, 1.33]	1.17 [0.96, 1.38]	1.14 [0.94, 1.34]
Controlled Motivation	–		*1.00 [0.86, 1.17]*	1.04 [0.89, 1.21]	*1.03 [0.89, 1.21]*	1.06 [0.92, 1.25]	1.05 [0.88, 1.22]	1.11 [0.94, 1.29]
Amotivation	–		**1.74 [1.49, 2.05]**	**1.67 [1.43, 1.95]**	**1.64 [1.41, 1.93]**	**1.59 [1.35, 1.86]**	**1.56 [1.30, 1.81]**	**1.47 [1.23, 1.70]**
Personality								
Openness to Experience	–		–		0.90 [0.78, 1.04]	0.95 [0.82, 1.09]	0.86 [0.73, 0.99]	0.92 [0.78, 1.06]
Conscientiousness	–		–		**0.53 [0.45, 0.62]**	**0.55 [0.47, 0.64]**	**0.60 [0.50, 0.71]**	**0.64 [0.54, 0.74]**
Extraversion	–		–		*1.01 [0.87, 1.17]*	0.81 [0.70, 0.94]	0.94 [0.80, 1.10]	0.77 [0.65, 0.87]
Agreeableness	–		–		0.88 [0.76, 1.02]	0.91 [0.78, 1.05]	0.82 [0.70, 0.95]	0.87 [0.75, 1.01]
Neuroticism	–		–		0.84 [0.72, 0.98]	0.93 [0.80, 1.08]	1.10 [0.91, 1.28]	1.18 [0.98, 1.37]
Covariates								
Age group								

Determinants	Null Model		Step 1		Step 2		Step 3	
	Start of Survey	End of Survey	Start of Survey	End of Survey	Start of Survey	End of Survey	Start of survey	End of survey
16 – 19 vs.	–		–		–		–	
20 – 25	–		–		–		1.24 [1.07, 1.46]	0.96 [0.81, 1.12]
26 +	–		–		–		1.29 [0.58, 2.16]	0.71 [0.31, 1.17]
Gender								
Men vs.								
Women	–		–		–		0.73 [0.60, 0.87]	**0.61 [0.50, 0.72]**
Other	–		–		–		1.65 [1.12, 2.22]	1.05 [0.70, 1.40]
Ethnicity								
White vs. BIPOC	–		–		–		**2.07 [1.76, 2.42]**	**1.86 [1.58, 2.17]**
General Contextual Effects								
Respondent variance	1.93 [1.60, 2.32]	1.78 [1.47, 2.13]	1.73 [1.41 – 2.09]	1.68 [1.39, 2.02]	1.66 [1.35, 2.02]	1.62 [1.32, 1.95]	1.54 [1.22, 1.91]	1.52 [1.22, 1.86]
PCV (%)	N/A		10.8	5.3	4.0	4.0	7.4	5.9
ICC	.371	.351	.345	.339	.339	.330	.319	.316
AUC	.932	.926	.928	.924	.923	.921	.914	.913
AUC change*	N/A		-.004	-.002	-.005	-.003	-.009	-.008
Bayesian R^2	.257	.250	.259	.253	.262	.255	.272	.269
Goodness of fit								
ELPD LOO (SE)	-6132.94 (41.59)		-6078.84 (43.16)		-6054.65 (44.06)		-5367.50 (44.74)	
Δ ELPD LOO	N/A		-54.64 (59.93)		-24.20 (61.67)		-687.15 (62.79)	
WAIC (SE)	12 132.89 (81.96)		12 042.66 (85.24)		12 000.00 (87.10)		10 645.22 (88.59)	
PPP	.507	.506	.504	.504	.508	.506	.499	.507

Note. Values denote Odds ratio and 89% credibility interval. GM = Global Motivation; RM = Research Motivation. PCV : proportional change in the variance. ICC: intra-class correlation coefficient. AUC: Area under the curve. ELPD LOO: expected log predictive density leave-one-out. PPP = Posterior predictive p-value. Δ = Change in estimate. *Values in bold a meaningful credible effect using the 89% HDI + ROPE (+/- 20%) decision rule. Values in *italic* are practically equivalent to 0 using the 89% HDI + ROPE (+/- 20%) decision rule. Research program 1 = Academic Resilience, Research program 2 = Cellphone use in Classroom, Research program 3 = Motivation and Well-being, Research program 4 = Physical Activity, Research program 5 = Pro-environmental behaviours. BIPOC = Black, Indigenous and People of Colour

Table 4.4

Stepwise Model Comparison of Effects

Contrasts	Start Survey			End Survey		
	Mean Δ	89% CrI	% ROPE	Mean Δ	89% CrI	% ROPE
			Δ Step 2 - Step 1			
Global Motivation						
Amotivation	0.04	[-0.18, 0.25]	84.9	0.03	[-0.18, 0.24]	87.1
Controlled	-0.03	[-0.21, 0.25]	81.8	0.06	[-0.17, 0.29]	79.5
Autonomous	-0.16	[-0.09, 0.39]	55.9	-0.16	[-0.40, 0.06]	56.6
Motivation for Research Participation						
Amotivation	-0.06	[-0.17, 0.28]	80.7	0.05	[-0.16, 0.28]	82.6
Controlled	0.03	[-0.24, 0.19]	85.8	-0.03	[-0.24, 0.18]	86.1
Autonomous	-0.06	[-0.19, 0.30]	77.2	-0.04	[-0.28, 0.19]	80.9
			Δ Step 3 - Step 2			
Global Motivation						
Amotivation	-0.02	[-0.20, 0.23]	85.9	0.04	[-0.25, 0.18]	85.0
Controlled	0.01	[-0.26, 0.23]	81.2	-0.01	[-0.25, 0.22]	81.7
Autonomous	0.01	[-0.25, 0.24]	79.2	0.05	[-0.29, 0.20]	78.3
Motivation for Research Participation						
Amotivation	-0.06	[-0.17, 0.28]	80.0	0.08	[-0.29, 0.15]	77.8
Controlled	0.03	[-0.25, 0.19]	84.3	-0.04	[-0.26, 0.18]	83.8
Autonomous	0.00	[-0.26, 0.24]	79.6	0.01	[-0.23, 0.25]	81.0
Personality						
Openness to experience	-0.05	[-0.27, 0.15]	84.5	-0.03	[-0.23, 0.18]	87.7
Conscientiousness	0.12	[-0.11, 0.35]	67.2	0.15	[-0.06, 0.39]	58.6
Agreeableness	-0.07	[-0.28, 0.15]	81.2	-0.04	[-0.25, 0.17]	85.6
Extraversion	-0.07	[-0.29, 0.14]	80.9	-0.05	[-0.26, 0.16]	84.3
Neuroticism	0.27	[0.05, 0.51]	25.6	0.23	[0.00, 0.45]	35.4

Note. Mean Δ = difference in estimated log odds. CrI = Credibility interval obtained from the highest density interval (HDI). % ROPE = proportion of the 89% HDI which lays within the ROPE

Table 4.5

Contrasts Across Hierarchical Levels of Motivation and Across Categorical Variables for Model 3

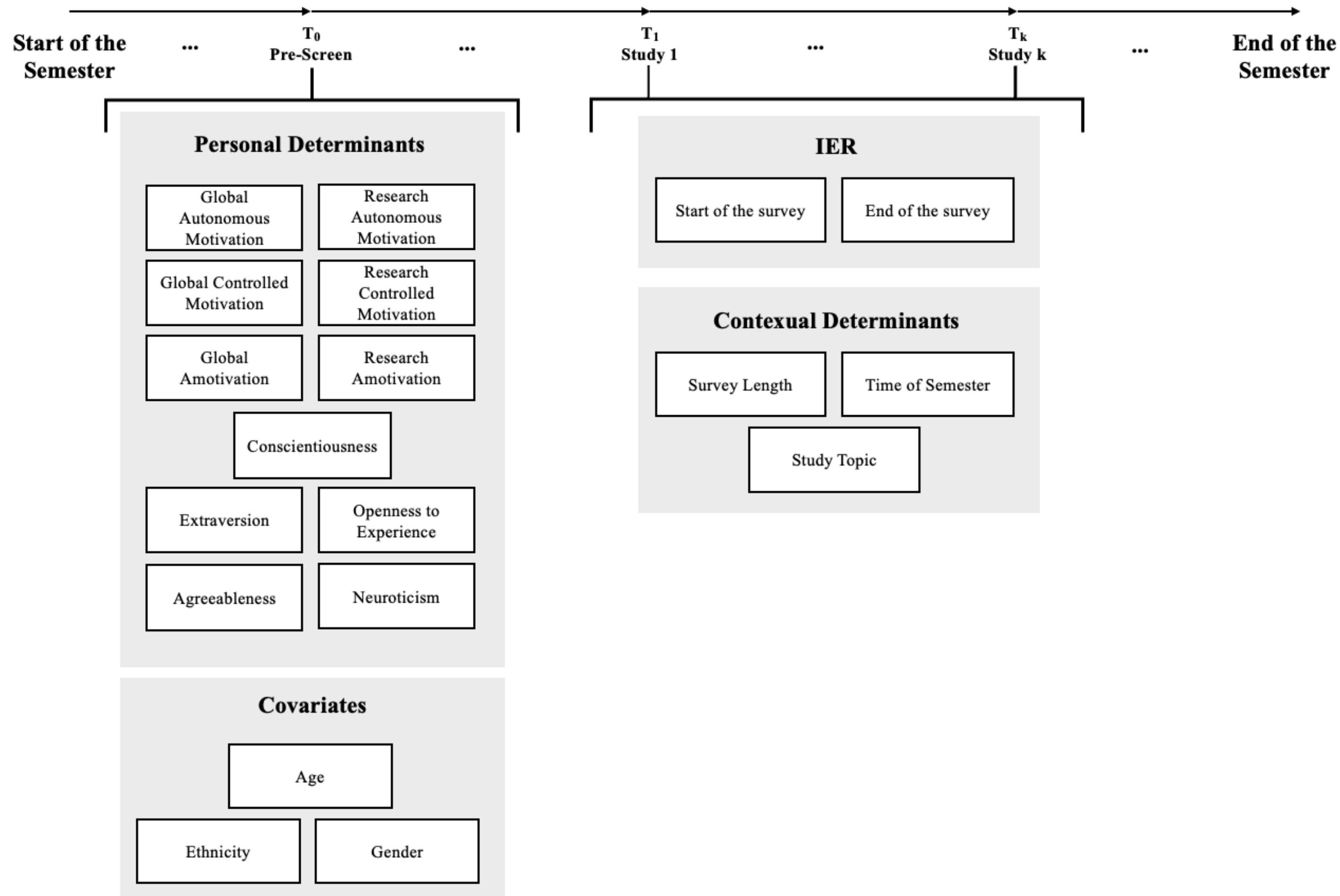
Contrasts	Start Survey			End Survey		
	Mean Δ	89% CrI	% ROPE	Mean Δ	89% CrI	% ROPE
Hierarchical Model of Motivation (Global – Research)						
Global Amotivation vs.						
Research Amotivation	-0.17	[-0.42, 0.06]	0.52	-0.27	[-0.49, -0.05]	0.21
Global CM vs.						
Research CM	0.21	[-0.02, 0.41]	0.41	0.01	[-0.19, 0.22]	0.95
Global AM vs.						
Research AM	0.23	[0.02, 0.43]	0.35	0.03	[-0.15, 0.22]	0.96
Study Topic (Within)						
Research program 1 vs.						
Research program 2	-0.03	[-0.71, 0.64]	37.9	0.20	[-0.44, 0.84]	35.9
Research program 3	-0.43	[-0.96, -0.09]	19.2	0.04	[-0.54, 0.45]	50.7
Research program 4	-0.28	[-0.72, 0.15]	34.1	-0.05	[-0.46, 0.37]	57.6
Research program 5	0.00	[-0.31, 0.31]	72.9	0.03	[-0.26, 0.33]	73.9
Research program 2 vs.						
Research program 3	-0.40	[-0.84, 0.02]	17.1	-0.16	[-0.56, 0.24]	50.4
Research program 4	-0.26	[-0.79, 0.28]	35.9	-0.25	[-0.74, 0.27]	37.4
Research program 5	0.02	[-0.59, 0.70]	40.0	-0.17	[-0.78, 0.43]	37.6
Research program 3 vs.						
Research program 4	0.15	[-0.40, 0.70]	41.6	-0.09	[-0.60, 0.44]	46.6
Research program 5	0.42	[-0.09, 0.94]	19.9	-0.01	[-0.50, 0.48]	50.9
Research program 4 vs.						
Research program 5	0.28	[-0.07, 0.61]	31.1	0.08	[-0.25, 0.39]	67.7
Study Topic (Between)						
Research program 1 vs.						
Research program 2	0.87	[0.04, 1.71]	4.0	0.68	[-0.13, 1.51]	11.9
Research program 3	-0.15	[-0.84, 0.53]	34.7	0.26	[-0.43, 0.95]	30.8
Research program 4	0.59	[-0.19, 1.36]	15.3	0.10	[-0.65, 0.86]	33.1
Research program 5	-0.24	[-0.88, 0.39]	33.6	0.53	[-0.10, 1.16]	15.5

Contrasts	Start Survey			End Survey		
	Mean Δ	89% CrI	% ROPE	Mean Δ	89% CrI	% ROPE
Research program 2 vs.						
Research program 3	-1.03	[-2.05, -0.03]	4.0	-0.42	[-1.40, 0.61]	20.7
Research program 4	-0.28	[-0.89, 0.36]	31.3	-0.58	[-1.17, 0.06]	11.7
Research program 5	-1.11*	[-1.77, -0.46]	0.0	-0.15	[-0.80, 0.51]	36.0
Research program 3 vs.						
Research program 4	0.74	[0.11, 1.32]	2.4	-0.16	[-0.76, 0.42]	39.4
Research program 5	-0.08	[-0.68, 0.51]	40.9	0.27	[-0.31, 0.87]	32.8
Research program 4 vs.						
Research program 5	-0.82*	[-1.35, -0.33]	0.0	0.43	[-0.07, 0.91]	18.4
Gender						
Men						
Women	0.32	[0.13, 0.51]	7.9	0.50*	[0.32, 0.68]	0.0
Other	-0.50	[-0.16, -0.84]	1.2	-0.04	[-0.38, 0.30]	68.3
Women						
Other	-0.82*	[-1.14, -0.50]	0.0	-0.54*	[-0.86, -0.22]	0.0
Age Groups						
16 to 19 years						
20 to 25 years	-0.21	[-0.39, -0.04]	36.9	0.04	[-0.12, 0.20]	97.1
26 years and older	-0.25	[-0.86, 0.37]	32.6	0.35	[-0.26, 0.99]	29.2
20 to 25 years						
26 years and older	-0.04	[-0.66, 0.60]	40.0	0.31	[-0.29, 0.97]	31.3

Note. Mean Δ = difference in estimated log odds. CrI = Credibility interval obtained from the highest density interval (HDI). % ROPE = proportion of the 89% HDI which lays within the ROPE. *Denotes meaningful differences (i.e., instances where the 89% HDI does not overlap with the ROPE). AM = Autonomous motivation. CM = Controlled motivation. Research program 1 = Academic Resilience, Research program 2 = Cellphone use in Classroom, Research program 3 = Motivation and Well-being, Research program 4 = Physical Activity, Research program 5 = Pro-environmental behaviours.

Figure 4.1

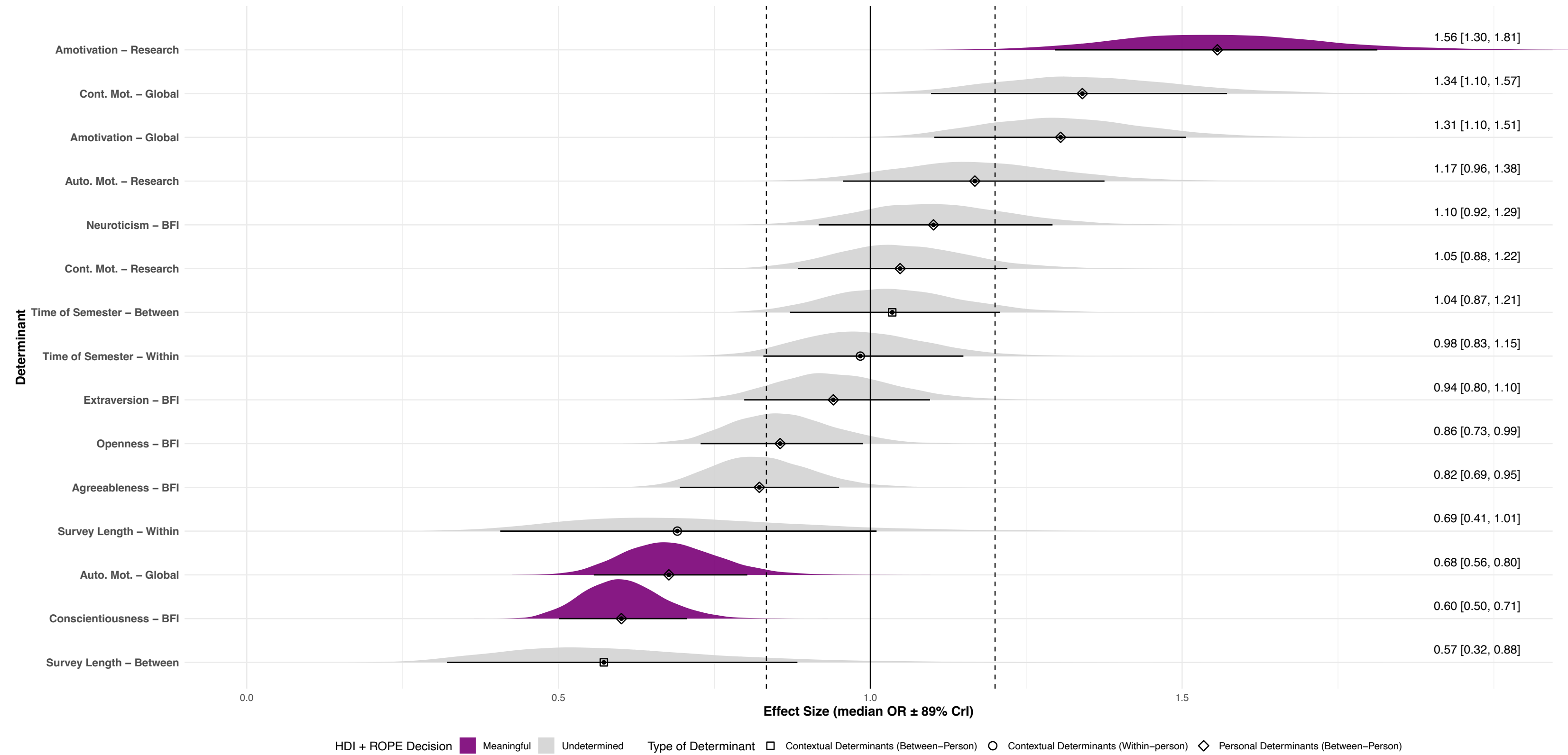
Data Collection Process



Note. k represents the number of studies completed by the participant and can take a value contained between two and seven.

Figure 4.2

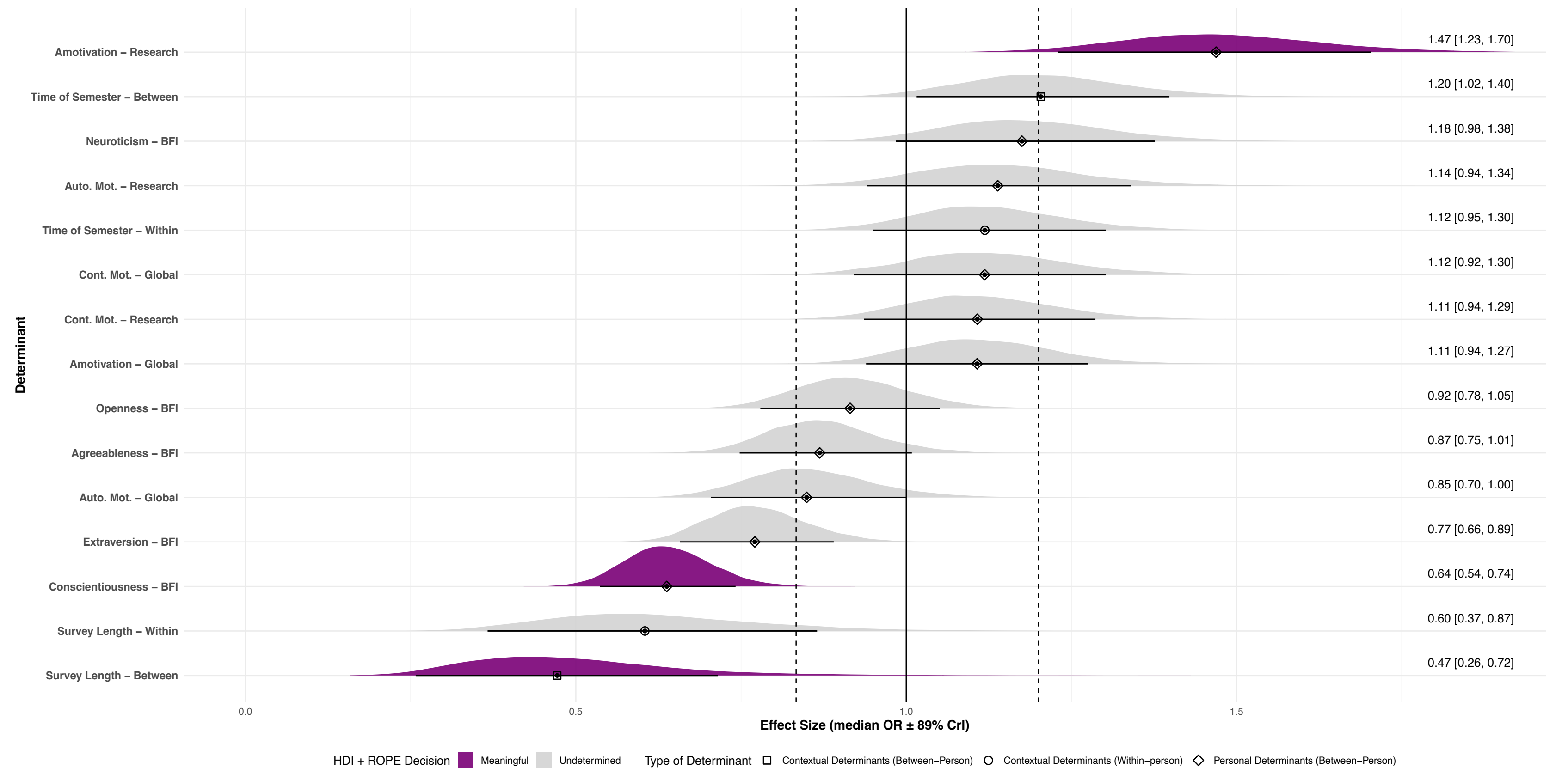
Start of the Survey - Effect of the Numerical Contextual and Personal Determinants on Odds of IER



Note. Aut. Mot. = Autonomous motivation. Cont. Mot.= Controlled motivation. BFI = Big five inventory. Values denote the median odds ratio from the posterior distribution. CrI = Credibility interval computed using the Highest Density Interval. The region within the dotted vertical lines denotes the Region of Practical Equivalence (ROPE). Distributions in light grey denote a non-meaningful effect (overlap of the 89% CrI and the ROPE). Distributions in purple denote a meaningful effect (no overlap of the 89% CrI and the ROPE

Figure 4.3

End of the Survey - Effect of the Numerical Contextual and Personal Determinants on Odds of IER



Note. Aut. Mot. = Autonomous motivation. Cont. Mot.= Controlled motivation. BFI = Big five inventory. Values denote the median odds ratio from the posterior distribution. CrI = Credibility interval computed using the Highest Density Interval. The region within the dotted vertical lines denotes the Region of Practical Equivalence (ROPE). Distributions in light grey denote a non-meaningful effect (overlap of the 89% CrI and the ROPE). Distributions in purple denote a meaningful effect (no overlap of the 89% CrI and the ROPE).

CHAPTER V : General Discussion

Summary of Thesis Objectives

Insufficient effort responding (IER) represents one of the biggest threats to data quality and research integrity faced by researchers from survey-reliant disciplines. Improving our ability to identify IER participants and furthering our understanding of the determinants promoting these egregious responses is thus of great relevance. The development of such knowledge is a necessary step for ensuring the quality of data and, consequently, the validity of survey research conclusions. The present thesis sought to contribute to the literature by introducing and validating a tool capable of improving the detection of IER and by examining the contextual and personal determinants of IER. In doing so, the overarching objective of this thesis was to generate tools and knowledge that can be applied to improve the quality and validity of conclusions drawn from survey research.

It is with this end in mind that the multidimensional IER detection approach (mIERda) was conceptualized. The mIERda is a robust and standardizable IER detection approach built on an innovative application of an unsupervised learning algorithm (i.e., fair-cut forest). This thesis incrementally demonstrates the validity of the mIERda, and positions it as an accurate and performant IER detection tool. After providing evidence of the mIERda's validity, the role played by contextual (i.e., survey length, time of the semester, study topic), and personal determinants (i.e., research participation motivation, global motivation, personality) on participants engagement in IER throughout the survey was investigated. In doing so, this thesis sheds light on the determinants of IER and highlights the differential motivational dynamics underpinning IER at the start and the end of the survey. This thesis provides evidence that amotivation for research participation increases participants' odds of engaging in IER throughout

the survey, whereas global autonomous motivation solely decreases participants' odds of engaging in IER at the onset of the survey. Results highlight why some individuals might be more likely of engaging in research attentively, and why others might be more prone to engage in IER.

Overall, this thesis achieved its overarching objective by presenting a series of six studies across three manuscripts. In the following sections, a summary of the main findings of these manuscripts will be provided. Afterwards, the implications, limitations, and future directions for research of this thesis will be discussed.

Summary of Main Findings

The main findings and implications of each manuscript included in the present thesis are summarized below.

Manuscript 1 - Introducing and Testing the Accuracy of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

Manuscript 1 sought to address limitations tied to conventional IER detection approaches by introducing a novel approach to IER detection, the IER detection approach (mIERda). The mIERda proposes a novel approach where IER indicators can be interpreted simultaneously by deferring the interpretation of these indicators to an unsupervised learning algorithm. Through four simulation studies, the objective was to provide initial evidence of the criterion validity of the mIERda by assessing its classification accuracy across a wide range of conditions (i.e., scales, prevalence, severity of IER, types of IER response patterns, simulated vs. human attentive data) and by comparing it to the current IER detection gold standard (i.e., sequential approach; a serial screening process where IER indicators are interpreted individually after establishing cutoffs for each indicator).

Differential Performance by IER Response Patterns. In Study 1a and 1b, it was observed that the performance of the mIERda differed by response patterns. For instance, it was noticed that the balanced accuracy of the mIERda decreased more strongly when the prevalence rate increased if the data solely contained bimodal or uniform IER respondents than when it was contaminated by a mixture of IER respondents. More precisely, results from Study 1a and 1b support that the mIERda might perform best when datasets contain multiple types of IER response patterns. Such findings align with the idea that the algorithm underpinning the mIERda seeks to identify outlying instances (Liu et al., 2008). When these "outlying" instances are overtly similar and prevalent, it is unsurprising for the mIERda to decrease in accuracy. Although the criterion validity of the mIERda was assessed in datasets containing individual response patterns, the mixed-respondents condition was considered as a more realistic representation of the criterion validity the mIERda in a real-life setting. Consequently, the following discussion centers on its performance in those instances.

Effect of survey length. Results from Study 1a and 1b revealed that the mIERda was more accurate when used to identify IER in longer scales (i.e., 240 items) vs. when used in shorter scales (i.e., 36 items). Similarly, results from Study 3 revealed that the balanced accuracy of the mIERda improved when two shorter scales were simultaneously screened for IER. This aligns with past literature reporting better performance when IER indicators are used in longer scales (Dupuis et al., 2018; Falk et al., 2025; Hong et al., 2020). Accordingly, results suggest that the mIERda might benefit from being used in longer scales or using more items.

Effect of severity of IER. Results from each study in *Manuscript 1* revealed that the classification accuracy of the mIERda was impacted by the severity of IER. It was observed that the mIERda was more accurate when the simulated IER respondents consistently provided IER

responses to all the items (i.e., fully IER), compared to when they were initially attentive and eventually shifted toward IER (i.e., partially IER). Once again, such findings align with the idea that as the quantity of information increases (i.e., relative proportion of responses provided as IER), measurement error decreases (Cronbach, 1951; Nimon et al., 2012).

Effect of prevalence rate. Results from Studies 1a, 1b, 2 and 3 suggest that when the data contained a mixture of IER respondents, the balanced accuracy of the mIERda remained relatively stable across prevalence rates. Although balanced accuracy was found to be stable across prevalence rates, results from Study 3 suggest that the specificity and sensitivity of the mIERda differed slightly across prevalence rate. The mIERda was found to be less sensitive and more specific when the prevalence of IER increased. Conversely, the mIERda was more specific and less sensitive when the prevalence of IER decreased.

The mIERda as a replacement of the sequential approach to IER detection. Results from Study 1a and 2 suggest that when a mixture of IER respondents was simulated, the mIERda outperformed, across conditions, the sequential approach in terms of balanced accuracy. This effect was observed in simulated and human-generated attentive data, regardless of sample size, severity of IER, prevalence rate, or scale. This implies that the mIERda correctly classified more respondents, irrespective of their true status (i.e., attentive vs IER), than the sequential approach. This positions the mIERda as a possible alternative to the current gold standard for IER detection (i.e., sequential approach).

Manuscript 2 - A Multi-Study Examination of the Nomological Validity of the Multidimensional Insufficient Effort Responding Detection Approach (mIERda)

The main purpose of *Manuscript 2* was to provide further evidence of the validity of the mIERda. While *Manuscript 1* introduced and provided evidence of the criterion validity of the

mIERda, its construct validity remained unexplored. *Manuscript 2* thus builds on *Manuscript 1* by examining the construct validity of the mIERda within its broader nomological network. To achieve this objective, *Manuscript 2* presents the results from a study examining the convergent, concurrent, discriminant, known-groups, and predictive validity of the mIERda using data arising from 18 distinct datasets ($N = 5014$) and five distinct research programs. The correlation between each validation variable and the mIERda was firstly computed at a dataset-level. A pooled correlation estimate was then obtained by pooling the correlation estimates from each dataset. In doing so, the goal was to obtain a stable, robust, and generalizable estimation of the construct validity of the mIERda.

Construct validity. The convergent validity of the mIERda was supported by the association observed between the mIERda and indicators of IER, which are not included in the mIERda. More precisely, the convergent validity was supported by the positive correlation between attention checks and self-reported measures of IER (i.e., self-reported effort, self-reported attention, self-reported data inclusion), as well as the negative correlation between the mIERda and completion time. (i.e., indicators of IER which are not included in the mIERda). Although the strength of the correlation between the mIERda and these indicators varied in strength, spanning from a moderate-to-strong correlation with attention checks and self-reported data inclusion to a weak correlation with completion time, findings aligned with the literature (Meade & Craig, 2012) and were in the expected direction. Furthermore, sensitivity analyses revealed that the mIERda identified most of the participants who failed more than one attention checks. The mIERda also identified as IER most participants who self-reported that their data should be excluded due to poor quality. Overall, the investigated indicators of IER were related, as expected, with the mIERda at the start and the end of the survey, which supports the

convergent validity of the mIERda. The discriminant validity of the mIERda was exemplified by its non-significant correlation with social desirability, global controlled motivation and controlled motivation towards research participation which aligned with results from past research reporting null association between IER and these variables (Cheung, 2019; Grau et al., 2019; Meade & Craig, 2012). The concurrent validity of the mIERda was supported by the small positive correlation of the mIERda with amotivation at a global and domain-specific (i.e., research participation) level. The predictive validity of the mIERda was further exemplified by the small negative correlation of the mIERda with conscientiousness, global autonomous motivation, and research autonomous motivation, which aligned with results from past studies reporting a similar association between IER and these variables (Arias et al., 2020; Bowling et al., 2016; Cheung, 2019). Similarly, the concurrent validity and known-group validity of the mIERda was supported by the positive correlation of the mIERda with the time of the semester and gender (0 = Women, 1 = Men), which supported past research stating that the prevalence of IER increases as participants complete the survey later in the semester (Cheung, 2019; Witt et al., 2011), and in men (Braitman et al., 2025; Cheung, 2019; Schneider et al., 2018; Witt et al., 2011). The construct validity of the mIERda was stable throughout the survey (i.e., start vs. end of the survey).

Prevalence of IER. Examination of the descriptive statistics showed that around one in three (32.5%) research participants were IER at the start of the survey. At the end of the survey, the average prevalence of IER was around 35.7% across datasets. Results from the frequency analysis revealed that the prevalence of IER was significantly greater at the end than at the start of the survey. However, the effect size was small, suggesting a limited increase in prevalence throughout the survey.

These results are in line with previous studies reporting a prevalence of IER ranging between 3% and 50% (Francavilla et al., 2019; Maniaci & Rogge, 2014; Meade & Craig, 2012), which provides additional evidence of the validity of the mIERda. While findings from previous studies suggest that the prevalence of IER increases as participants progress across the survey (Bowling et al., 2021) and claim that IER results from participants' fatigue as participants progress through as survey (Krosnick, 1999; Oppenheimer et al., 2009). Results from *Manuscript 2* provide additional nuance to these results and suggest that the prevalence of IER is already approximately 33% within the first 50 items of the survey. and slightly increases at the end of the survey. Such findings demonstrate that IER might be more common at the start of the survey than initially believed (Ward & Meade, 2023). Although this divergence might be due to differences in operationalization of IER, results from *Manuscript 2* challenge the assumption that IER is uncommon at the start of the survey.

Taken together, findings from *Manuscript 2* support the construct validity of the mIERda to identify IER throughout the survey and bring nuance to past literature by providing evidence that the prevalence of IER is already relatively high at the start of the survey across numerous datasets arising from various research programs.

Manuscript 3 - Joint Exploration of the Contextual and Personal Determinants of Insufficient Effort Responding Over Time – A Self-Determination Theory Perspective

After conceptualizing the mIERda (*Manuscript 1*) and providing evidence of its criterion and construct validity (*Manuscript 2*), the main focus of *Manuscript 3* was to highlight the determinants of IER. To achieve this goal, I relied on a longitudinal prospective approach and analyzed data arising from 23 datasets and five research programs. In doing so, the determinants

of IER of 2216 participants were analyzed over time using a stepwise joint multilevel Bayesian logistic regression model

Contribution of motivation. The results from *Manuscript 3* suggest a differential contribution of motivation that depended on the degree of internalization (i.e., autonomy) and specificity. At a global level, results suggested that being with high levels of global autonomous motivation might have a protective effect against IER at the start of the survey, but that this protective effect wanes as participants progress through the survey. Other types of global motivation did not meaningfully predict participants' engagement in IER throughout the survey. At a domain-specific level, participants' amotivation for research participation was tied to increased odds of engaging in IER throughout the survey. Contrastingly, autonomous and controlled motivation for research participation did not impact one's propensity of engaging in IER at any point throughout the survey. Overall, the meaningful effect of amotivation at a domain-specific level aligns with results from past research (Cheung, 2019) and corroborates the claim that IER results from a lack of motivation to read, interpret, and respond to survey items as posited by Huang and colleagues (2012). Such findings suggest that researchers interested in decreasing the prevalence of IER in their studies might benefit from implementing interventions seeking to decrease participants' amotivation towards research participation.

The role of personality. Participants' level of conscientiousness had a protective effect on subsequent engagement in IER. More precisely, participants with greater levels of conscientiousness were significantly less likely to engage in IER throughout the survey (i.e., at the start and end of the survey). This corroborates previous research (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017; Ward & Meade, 2023). Contrary to what had been hypothesized, participants' agreeableness, openness to others, extraversion, and emotional stability did not

meaningfully decrease the odds of IER. The effects were, however, in the expected direction, which partially supports past research on this topic (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017; Ward & Meade, 2023).

Effect of survey length on IER. Contextual factors had little effect at a within-person level. The length of the survey is no exception. In other words, completing a longer-than-usual survey was not associated with an increased propensity of engaging in IER throughout the survey. However, participants who tended to complete longer studies were also less likely to engage in IER at the end of the survey than those who generally completed shorter studies. This effect corroborates the negative association between survey length and IER at the end of the survey observed in *Manuscript 2*. Although unintuitive, these results suggest that the association between survey length and IER might rather reflect inter-individual differences in one's general willingness to partake in longer studies. These findings bring nuance to the literature stating that IER is more common in longer surveys (Berry et al., 1992; Brower, 2018; Herzog & Bachman, 1981). The difference in results between the present research and the literature might be due to the lack of an experimental design. Contrary to these studies, participants in the present research were not randomly assigned to a condition and got to decide and choose which study they participated in. It is thus possible that participants who are disinterested by research, rushed, or disengaged tend to sign up for shorter studies in an attempt to obtain their compensation as quickly and as effortlessly as possible.

Effect of the time of the semester. Results from *Manuscript 3* imply that when participants completed a survey later than usual within the semester, it did not meaningfully impact their odds of engaging in IER throughout the survey. While results revealed that participants who tended to complete surveys later in the semester *might* be more likely to engage

in IER at the end of the survey than those who participated earlier in the semester, the effect lacked meaningfulness. This nuances past literature stating that participants who participate later in the semester are more likely to engage in IER (Cheung, 2019; de Calvo & Reich, 2007; Witt et al., 2011) and suggests that this effect might not be as meaningful as initially believed and might be better explained by other variables. It nonetheless remains possible that the odds of IER do not differ whether participants participate, on average, in research between the 2nd and the 9th week of the semester, but that a difference could be observable when a participant completes all their studies at the end of the semester. It would thus be interesting for future research to examine such dynamics and whether this effect differs when we treat the time of the semester categorically (e.g., start of the semester, end of the semester, last week of the semester).

Effect of study topics. Findings from *Manuscript 3* suggest that the odds of engaging in IER, for a given person, did not differ across study topics. In other words, for a participant who is not inherently prone to engaging in IER, the study topic does not seem to have a significant effect on their odds of IER. This idea is coherent with past studies suggesting that interest in a study topic is not associated with engagement in IER (Hasselhorn et al., 2025). A few between-person differences in odds of engaging in IER at the start of the survey were, however, identified across study topics. Interestingly, these differences were not observed at the end of the survey. This suggests that some study topics might be more attractive to individuals who are more likely to engage in IER at the start of the survey.

Implications of the Present Thesis

The broader goal of this thesis was to contribute to the literature pertaining to IER. More precisely, the goal was not only to improve researchers' ability to identify IER in survey research data, but also to deepen our understanding of the determinants underpinning IER. The thesis

achieves this goal by proposing and validating a novel approach to measure this phenomenon and by presenting a robust investigation of the effect of various contextual determinants such as survey length, and individual determinants, such as personality traits and motivation at a global or context-specific (i.e., research participation) level. These implications are discussed below in greater detail.

Theoretical Implications

From a theoretical standpoint, this thesis contributes to the field of Self-Determination Theory (SDT) by demonstrating that individuals' motivational orientations, at a global and research participation level, modulate the quality of the data provided by participants partaking in survey research. Although past research has demonstrated that individuals with greater levels of amotivation are more likely to engage in IER (Cheung, 2019), this thesis provides additional nuance by not only supporting that research amotivation is associated to IER; but also, by demonstrating that amotivation for research participation increases participants' odds of engaging in IER throughout the survey. The present thesis demonstrates that participants with higher levels of global autonomous motivation are more likely to be attentive at the start of the survey. Taken together, these results suggest that motivation, as defined by SDT, could be a meaningful determinant of participants' response quality across studies of various topics, even when accounting for the effect of known contextual and personal determinants, and other covariates associated with IER.

While additional research is needed to thoroughly assess and shed light on the intricacies underpinning the effect of motivation on IER, *Manuscript 3* provides empirical evidence supporting the claim that IER is linked to a lack of motivation to read, interpret, and respond to survey items as defined by Huang and colleagues (2012). In fact, the "lack of motivation" which

is central to the definition of IER aligns with SDT's definition of amotivation. Amotivation describes a state in which the individual faces the absence of motivation to engage in a behaviour or in an activity; the individual lacks purpose, meaning or perceived value as they engage in the behaviour (Ryan & Deci, 2017). It is broadly defined as the absence of motivation. SDT states amotivation can arise from two main sources. The individual can become amotivated (a) in the face of a feeling of perceived incompetence or if they feel like the behaviour will not yield the expected outcome (Pelletier, Dion, Tuson, & Green-Demers, 1999), or (b) in the face of indifference about the activity and its associated outcomes (Ryan & Deci, 2006; Vansteenkiste, Lens, et al., 2004). The former is especially relevant to consider as we investigate the determinants of IER. Interestingly, research supports that amotivation arising from indifference can stem from two sources: (1) either the individual was never exposed to the value of acting (Ryan & Deci, 2017), or (2) the individual determine that they have no intrinsic interest in the behaviour nor for the outcome, thus autonomously minimize or withdraw their engagement (Ryan & Deci, 2006; Vansteenkiste, Lens, et al., 2004). In the former instance, this would refer to participants who engage in IER because they were never shown or explained the value of expending a sufficient effort responding while participating in survey research. In contrast, in the latter instance, the person might decide to engage in IER to allocate spared resources and their energy to things and activities that are deemed as more valuable to them. For them, getting their compensation as effortlessly as possible, by engaging in IER, might be the best way of channeling their energy towards their goals while minimizing their engagement in "futile" activities.

Considering the empirical evidence found in *Manuscript 3,* suggesting that amotivation for research plays a meaningful role in predicting IER, the Pelletier and Rocchi (2023)

framework also provides an interesting line of inquiry for future research. The authors propose that different types of nonengagement are characterized by specific underlying need issues (e.g., competence frustration) and, as such, effective intervention strategies must be tailored to each scenario. In the context of IER, this suggests that autonomy-supportive interventions highlighting the relevance, importance, and value of participating in research might decrease odds of engaging in IER for some individuals (e.g., reduce pressure for introjected nonengagement). Future research could strive to distinguish various types of nonengagement in IER behaviours to establish more tailored interventions. Nonetheless, this thesis provides us with insight pertaining to the relevance of considering and targeting amotivation for research participation in a context of IER, but also about the possible limitations of such interventions.

The overall absence of meaningful effect of controlled motivation (i.e., research participation and global level) means that alternative forms of “controlling” interventions (e.g., rendering the compensation dependent on the quality of the responses provided) should not directly increase the prevalence of IER. As demonstrated by the results from the present thesis and from past research (Cheung, 2019), being predominantly driven by external (i.e., “to obtain the compensation”) or introjected (i.e., “to avoid guilt”) motives, at a global or a research participation level, does not result in more IER. This implies that interventions which aim to increase the salience of personal goals through a controlled approach although suboptimal, could be effective. While such interventions were not tested as it was outside the scope of the present thesis, this reasoning is in line with past research reporting the effectiveness of interventions relying on warning messages (i.e., warning that data would be screened for IER and that engaging in IER would result in loss of compensation; Huang et al., 2012). While such interventions are controlling in nature and should thus be ineffective in increasing engagement;

additional mechanisms might be at play. On the one hand, controlling interventions are often portrayed derogatively due to their negative association with wellbeing and their pervasive impact on the internalization of behaviours (Ryan & Deci, 2000, 2017). On the other hand, while controlling interventions would not be a viable option if the objective was to ensure the perennity of the behavioural change; they remain a useful tool to bring individuals to temporarily engage in a behaviour (Deci, 1971). As such, these warnings might be effective since they highlight the conditional nature of the compensation. In other words, this nuances the contingency of the reward: suddenly it is not solely a question of responding (i.e., completion-contingent reward), but rather a question of responding well (i.e., performance-contingent reward; Ryan et al., 1983). Framing the incentive in such a way could either foster a greater level of engagement by decreasing the perceived distance with the research team (i.e., by increasing relatedness), by conveying competence feedback (i.e., by increasing competence), or by increasing the salience of values that might be important to the individual such as diligence and persistence. However, it is plausible that these interventions do not truly reinforce engagement and solely appear to be effective since they prompt participants to engage in appearance-management (i.e., by providing responses that *look* attentive rather than attempting to *be* attentive). Future research would benefit from examining the impact of such interventions on participants' motivation, especially in the long-term.

Nonetheless, it remains likely that the lack of association between IER and controlled motivation is tied to the context in which the research takes place. More precisely, in a university-based subject recruitment pool, each participant is driven, to a certain extent, by the obtention of their compensation (i.e., course credit). If participants were not motivated to a certain extent by their compensation, they would not have attempted to obtain it. In such

instances, it might be more interesting to focus on what the participants are trying to achieve beyond said compensation.

Hierarchical Model of Intrinsic and Extrinsic Motivation. The results from *Manuscript 3* offered partial support to the hierarchical model of intrinsic and extrinsic motivation (Vallerand, 1997). While the various forms of global motivation were positively correlated to their context-specific counterpart (i.e., research participation), results failed to provide evidence that motivation at a global level had a weaker effect on IER than motivation at a domain-specific level (i.e., research participation). A stark example of this finding is the fact that global autonomous motivation was a negative and meaningful determinant of IER while the effect of autonomous motivation for research participation (i.e., domain-specific) was not found to be meaningful. While the present research assessed whether the effect of motivation differed by level of specificity, it did not formally assess whether global motivation had a top-down effect on motivation for research participation nor if motivation for research participation had a bottom-up effect on global motivation.

Nonetheless, this thesis sheds light on interesting distinctions in the role of motivation depending on the level of specificity. While amotivation for research participation was a positive predictor of IER at the start and the end of the survey; the effect of global amotivation was deemed non-meaningful. Conversely, while autonomous motivation for research participation failed to meaningfully predict odds of IER; global autonomous motivation was found to decrease the odds of IER, but solely at the start of the survey. Such finding might sound unintuitive given that motivation at a global level denotes individual tendencies and should thus be relatively stable throughout the survey (Vallerand, 1997). While more research is needed to understand this effect, we can posit that global and domain-specific motivation acts at two distinct levels. Since

autonomous global motivation influences the quality of participants' motivation as they apprehend various tasks or activities, it might have a stronger role to play in shaping how participants approach the start of the survey. It is, however, likely that as participants progress through the survey, the specificity of the tasks increases which could explain why the role of global motivation vanishes. From an intervention standpoint, distinguishing the effect of global and domain-specific motivation is critical. While it might be difficult to change one's motivation at a global level (i.e., disposition), given its trait-like nature, domain-specific motivation can be more easily targeted through interventions seeking to lower levels of amotivation.

Practical Implications

Beyond the several theoretical implications presented above, results from the present thesis have numerous practical implications for researchers and data analysts in the field of social psychology and from any discipline relying on survey data. More precisely, the present thesis has multiple relevant implications for researchers interested in identifying IER and decreasing its occurrence in their samples.

IER detection. The results emerging from the present thesis have practical implications for researchers interested in identifying IER. The introduction and validation of the multidimensional IER detection approach (mIERda) is central to these implications. By incrementally examining its criterion and construct validity, empirical evidence arising from the manuscripts contained herein position the mIERda as a novel, standardizable, and valid approach to IER detection. Whereas the current gold standard IER detection approach (i.e., sequential approach) faces limitations due to its reliance on cutoffs and its non-standardizable nature, this thesis offers a novel perspective to IER screening and introduces a way to move beyond these

limitations. The mIERda achieves this goal by relying on an innovative adaptation of an unsupervised machine learning algorithm to enhance and facilitate the detection of IER.

As demonstrated in *Manuscript 1*, the mIERda was found to have a greater level of balanced accuracy than the sequential approach. Such a result implies that the mIERda outperformed the current gold standard approach in terms of correctly classifying attentive and IER participants. Beyond providing evidence of its criterion validity, such result implies that the mIERda might constitute a tangible way to improve the accuracy of IER screening performed by researchers which positions it as a potentially suitable replacement to the sequential approach. Given the known pervasive impact of IER (e.g., Arias et al., 2020, 2022; Huang, Liu, & Bowling, 2015; Kam & Meyer, 2015; Oppenheimer et al., 2009), the results from the present thesis also imply that by increasing the accuracy of the IER screening performed by researchers, the mIERda might represent a way to improve the validity of conclusions drawn from survey research data. More research is however needed to explore this matter.

Estimation of sample size and statistical power. Results regarding the prevalence of IER have implications regarding the recruitment of participants and estimation of the required sample size. More specifically, results from *Manuscript 2* and *Manuscript 3* indicate that approximately 1 in 3 research participants seems to engage in IER at the start of the survey and approximately 2 in 5 participants appear to be engaging in IER at the end of the survey. Furthermore, results from *Manuscript 2* suggest that approximately 50% of participants engaged in IER at any point throughout the survey. To replace these participants, this implies that researchers should recruit around 30 to 50% more participants than required in order to preserve a sufficient sample size and maintain statistical power after performing IER screening. Such findings also have financial implications for researchers, as recruiting more participants is

necessarily tied with additional expenses (i.e., compensation, extra survey platform fees). Despite the known pervasive impact of IER on data quality, such result might deter researchers from identifying and addressing the presence of IER in their data. To limit the consequences of IER, researchers should plan for this eventuality by dedicating additional funds to participants' recruitment when applying for research grants.

Decreasing amotivation for research participation. In addition to providing a novel tool to facilitate the detection of IER respondents, the present thesis suggests researchers could potentially decrease the prevalence of IER by creating a research context which decreases participants' amotivation. This could be done by providing rationale, by explaining why their participation and attention matters, and by helping participants see the value and relevance of the study. Doing so might increase their willingness to remain attentive throughout the survey and decrease IER by fostering greater levels of autonomous motivation, but most importantly, by decreasing their level of amotivation towards research.

Data collection period within the semester. Results from *Manuscript 3* suggest that the time of the semester had a limited effect (i.e., non-meaningful) on participants engagement in IER. While more research is needed to examine this effect at a more granular level (e.g., comparing the last day of the semester to the remainder of the semester), this suggests that it might not be necessary for researchers concerned about the presence of IER in their data to limit their data collection period, within a given semester, when recruiting participants from a subject pool.

Shorter surveys? Yes and no. The results from *Manuscript 3* suggest that participants engaging in IER at the end of the survey also tend to sign up for shorter studies. Additionally, as participants completed longer studies, they were not meaningfully more likely to engage in IER

at the start nor the end of the survey. As such, while avoiding excessively long surveys is always good practice, researchers should remain aware that shorter studies might make them easier targets for IER respondents. While more research is needed, it might be best for researchers to ensure their studies look slightly demanding and engaging, all while remaining short, to avoid attracting participants who might be more prone to engaging in IER. Researchers, for example, might benefit from dividing longer surveys into shorter components, and administering each component of the survey in the context of a multi-part study. This method could provide a way to generate a research setting that requires a specific level of engagement, rendering the study unattractive to disengaged participants, while preventing participants' exhaustion during study completion. More research is, however, required to establish the suitability of such an approach.

Limitations of the Thesis and Directions for Future Research

While the present thesis offers compelling evidence of the performance of the mIERda and a thorough investigation of the determinants of IER, room for improvement remains. These limitations and how they can be addressed by future research are discussed below.

Improving the mIERda

While this thesis introduced and provided preliminary evidence of the criterion validity of the mIERda, certain limitations remain. In a first instance, in *Manuscript 1*, the mIERda was solely examined using datasets containing 500, 1000 and 2000 participants. While the classification accuracy of the mIERda was found to be relatively stable across these sample sizes, future research would benefit from examining its classification accuracy in smaller and larger samples. More research is needed to understand whether the presence of negatively worded items and inter-factorial correlation impact the performance of the mIERda. Future research would also benefit from investigating whether the mIERda can be optimized either by

manipulating the algorithm's hyperparameters and the algorithm used to establish the mIERda's cutoff. Additionally, more research is needed to examine whether adding or removing indicators of IER can improve the classification accuracy of the mIERda. Although the list of IER indicators currently included in the mIERda were selected for their demonstrated performance and applicability to a broad range of research designs, future research would benefit from investigating whether this list could be optimized.

Effect of IER on Data Quality

While the present research has assessed the convergent validity of the mIERda and its accuracy, the actual improvement in data quality ensuing remains to be assessed. Quantifying the improvement in data quality was outside the scope of the present thesis. Future research should focus on assessing whether using the mIERda leads to differences in factorial structure, scale reliability, correlations between variables, statistical power, and effect sizes. This would provide further evidence of the validity and the contribution of the mIERda.

Observational Design

As discussed in *Manuscript 3,* survey length was unexpectedly associated with IER in a way where individuals who tended to complete longer surveys were also less likely to engage in IER at the end of the survey. This effect was starkly in opposition with the literature suggesting that the odds of IER increase at the end of longer surveys (Berry et al., 1992; Brower, 2018; Herzog & Bachman, 1981). This divergence could be due to the lack of experimental design and could result from a selection bias where participants who are more strongly interested in participating in research might also be more likely to participate in longer surveys. Such hypothesis is coherent with the weak, yet significant, negative correlation observed in *Manuscript 3* between participants' average survey length and amotivation (i.e., at a research

participation and global level), as well as the positive correlation observed between autonomous motivation (i.e., research participation and global level) and participants' average survey length. It would be of great interest to implement an experimental manipulation where participants are randomly attributed either to a short or long survey condition and to assess the impact of such manipulation across studies of various topics. This could be done while controlling for the time of the semester and other determinants of IER such as motivation at a global and research participation-level.

Methodological Representativity

While the criterion validity of the mIERda is assessed across multiple conditions (i.e., IER severity, prevalence, types of respondents, sample size and scales) in the present thesis, it remains unknown how the mIERda truly performs in other scales and contexts. Future research may benefit from assessing the performance of the mIERda other contexts (e.g., unidimensional scales, surveys with three or more scales, scales with varying number of Likert number points), and conditions (e.g., smaller prevalence rate, prevalence rates greater than 50%, intermittent IER respondents). The present thesis also solely assessed the performance of the mIERda in scales arising from the field of psychology; more research is required to assess whether results generalize to scales from other disciplines (e.g., economy, business, political science).

Sample Representativity

The datasets used to examine the determinants of IER in *Manuscript 3* were only made up of undergraduate students at the University of Ottawa. While the analyses accounted for various possibly confounding variables (i.e., gender, ethnicity, age); it remains that the generalizability of the results in a broader sample remains unknown. For instance, it is possible that the difference in likelihood of IER observed between men and women results from

participants' developmental stage rather than their gender identity. It is indeed known that men tend to remain more impulsive than women in their early adulthood. Nonetheless, the reported difference across gender corroborates results from previous research (Braitman et al., 2025; Schneider et al., 2018). More research is needed to understand why the prevalence of IER differs across genders. Such differences suggest that other mechanisms that are not captured by motivation, personality nor contextual determinants might be at play. Future research would benefit from investigating whether the determinants of IER remain the same in a more age-diverse sample. Similarly, in *Manuscript 2*, the construct validity of the mIERda was solely assessed using samples of undergraduate students. Future research would also benefit from examining the construct validity of the mIERda in a more diverse sample.

Examination of IER Through a Multicultural Lens

Although *Manuscript 3* accounted for the effect of ethnicity on subsequent engagement in IER, future research would benefit from examining why Black, Indigenous and People of Colour (BIPOC) participants were more likely to engage in IER than White participants. While such investigation was beyond the scope of the present thesis, it could be of interest to conduct a mixed-method study to determine *if* BIPOC participants are truly more likely to engage in IER and if so, *why* and *when* are they more likely of engaging in IER.

Such examination would be relevant especially since the results from *Manuscript 3* are coherent with past research reporting cross-cultural differences in odds of IER (Grau et al., 2019). Grau and colleagues (2019) hypothesized that participants from Western, educated, industrialized, rich and democratic (WEIRD) societies might feel more comfortable withdrawing or refusing to complete the survey, which would prevent them from engaging in IER in the first place. It could also be of interest to build on Social Identity Theory (Tajfel, 1978; Tajfel &

Turner, 1979). If BIPOC participants, as a minority group, perceive that surveys are designed by, for, and about the majority group; they may feel like the survey fails to reflect their own experiences. In such instances, participating in research might feel “useless” or “less meaningful”. Such feeling might translate as disengagement, hence as IER. While research in psychology is criticized for being excessively “WEIRD” from a participant and a researcher perspective (Henrich et al., 2010), BIPOC participants might perceive research as a White-centered endeavour which highlights their minority status. In such instances, research participation might feel like identity work, consequently increasing burden and disengagement. Future research would benefit from investigating whether interventions seeking to foster in-group belongingness in research could help decrease the differences in odds of IER observed between BIPOC and White participants. It could also be of interest to investigate how in-group cultural congruence and incongruence between participants and researchers influence the odds of engaging in IER.

Motivational Interventions

While the present research supports the idea that respondents’ global autonomous motivation and amotivation toward research participation impact their likelihood to engage or not in IER, the present research does not evaluate the effectiveness of manipulating motivation in decreasing engagement in IER. It could thus be of interest to design and test the effectiveness of motivational interventions aimed at decreasing levels of research amotivation on respondents’ engagement in IER. For instance, more research is needed comparing the effect of autonomy-supportive interventions on the odds of IER. A study could compare the prevalence of IER in participants in a control group, those exposed to warning messages (i.e., controlling condition), and those who are provided rationale for participating attentively in research and/or are explicitly

offered the option of quitting the survey at any point while receiving the compensation (i.e., autonomy-supportive condition). Such a design would help us determine whether autonomy-supportive interventions are more effective than controlling interventions in decreasing the prevalence of IER and if this effect differs as participants progress throughout the survey (i.e., at the start vs. at the end of the survey).

It would also be of interest to assess whether participants' needs satisfaction and frustration moderate the effect of such interventions, and to investigate whether such interventions have a bottom-up effect on participants' motivation at a global level. It is likely that participants with greater levels of needs frustration might benefit more strongly from an autonomy-supportive intervention (i.e., from an intervention that does not further frustrate their need for autonomy and rather supports it) which could translate into less IER. These participants might also experience greater frustration when exposed to controlling-interventions which could translate into more IER. Conversely, participants with greater levels of need satisfaction might be less sensitive to the controlling and autonomy-supportive nature of the intervention. In other words, both types of intervention could translate into less IER for these individuals. Such a distinction would be of relevance to understanding whether reliance on controlling interventions to decrease IER might be detrimental in some instances.

"Attack of the Clones": Bots vs. IER Detection

While the present thesis aimed to facilitate the identification of IER respondents and the determinants of IER, the accuracy of the mIERda as it regards the identification of bots (i.e., a program or script used to fill surveys with fake responses) remains unexplored. The expected performance of the mIERda in such instances depends on numerous factors, namely (a) the response generation distribution/process and (b) the prevalence of bots in the survey. If bots

generate responses under a uniform, bimodal or normal distribution, or any other repetitive response pattern, there is no indication suggesting that the mIERda would not be able to identify bots. In fact, since bots are technically "fully inattentive" (Falk et al., 2025), it is safe to assume that it is most likely easier for the mIERda to identify those, compared to IER participants. The rise of artificial intelligence and machine learning, however, brings new challenges. More "human-like" bots based on more sophisticated algorithms are starting to appear (Irish & Saba, 2023), and the accuracy of the mIERda and bots' detection approaches remains unknown under such conditions. Future research would benefit from investigating whether the mIERda could potentially be used for bots' detection.

Conclusion

This doctoral thesis provides research teams conducting survey research with tools and insight to detect IER in their sample. The mIERda was created to enable researchers with the ability to identify the IER hiding in their data. Doing so, the goal of this thesis was not only to contribute theoretically to the literature by robustly examining the determinants of IER, but also to provide an applied contribution by introducing and validating a tool which could potentially improve to the overall quality of the research conducted with surveys.

Such a contribution is especially warranted as we consider that ever since the first survey in the late 1800s, the context in which surveys are administered has changed dramatically. We moved from a face-to-face to a fully online-based administration. As the distance grew wider between researchers and participants, so did the distance between the act of participating and the greater purpose tied to this participation. The first survey conducted in the streets of London by Booth in the late 1800s led to profound societal changes. Such changes were solely possible since respondents provided a truthful and accurate depiction of their living conditions (Ornstein,

2013). Perhaps the reason the first surveys led to such significant societal changes can be tied to the fact that the people in the streets of London, for once, felt seen and heard in a system that made them feel invisible. Perhaps instead of solely focusing on *how* and *why IER happens,* we should ask ourselves *what is in it for the participant*? If the sole benefit of participating is the compensation, how can we truly expect participants to expand efforts beyond those needed to reap said benefit? Researchers might gain from closing the distance between them, the participants, and the benefits of their research. In a way, it might be time for researchers themselves to put forth a sufficient effort in addressing insufficient effort responding.

References

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alfons, A., & Welz, M. (2024). Open science perspectives on machine learning for the identification of careless responding: A new hope or phantom menace? *Social and Personality Psychology Compass*, *18*(2). https://doi.org/10.1111/spc3.12941

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D. *American Journal of Preventive Medicine*, *10*(2), 77–84. https://doi.org/10.1016/S0749-3797(18)30622-6

Andrews, P., & Meyer, R. G. (2003). Marlowe–Crowne Social Desirability Scale and short Form C: Forensic norms. *Journal of Clinical Psychology*, *59*(4), 483–492. https://doi.org/10.1002/jclp.10136

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*(6), 2489–2505. https://doi.org/10.3758/s13428-020-01401-8

Arias, V. B., Ponce, F. P., & Martínez-Molina, A. (2022). How a Few Inconsistent Respondents Can Confound the Structure of Personality Survey Data: An Example with the Core-Self Evaluations Scale. *European Journal of Psychological Assessment*. 39(6), 400-407. https://doi.org/10.1027/1015-5759/A000719

Ashley, M., & Shaughnessy, K. (2021). Predicting Insufficient Effort Responding: The Relation Between Negative Thoughts, Emotions, and Online Survey Responses. *Canadian Journal of Behavioural Science*, *55*(3). 1–13. https://doi.org/10.1037/cbs0000308

Ashton, M. C., Lee, K., & Goldberg, L. R. (2007). The IPIP–HEXACO scales: An alternative, public-domain measure of the personality constructs in the HEXACO model. *Personality and Individual Differences*, *42*(8), 1515–1526. https://doi.org/https://doi.org/10.1016/j.paid.2006.10.027

Aviv, A. L., Zelenski, J. M., Rallo, L., & Larsen, R. J. (2002). Who comes when: personality differences in early and later participation in a university subject pool. *Personality and Individual Differences*, *33*(3), 487–496. https://doi.org/10.1016/S0191-8869(01)00199-4

Barnes, T. (2018). *Determining Cutoffs for the Psychometric Synonyms Analysis To Detect IER* [Doctoral Thesis, Wright State University.]. Core Scholar. https://corescholar.libraries.wright.edu/etd_all/2013

Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology*, *61*(2), 226–244. https://doi.org/10.1037/0022-3514.61.2.226

Baxter, D., & Pelletier, L. G. (2020). The roles of motivation and goals on sustainable behaviour in a resource dilemma: A self-determination theory perspective. *Journal of Environmental Psychology*, *69*. https://doi.org/10.1016/j.jenvp.2020.101437

Beach, D. A. (1989). Identifying the Random Responder. *The Journal of Psychology*, *123*(1), 101–103. https://doi.org/10.1080/00223980.1989.10542966

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*, *4*(3), 340–345. https://doi.org/10.1037/1040-3590.4.3.340

Bowling, N. A., Gibson, A. M., & Desimone, J. A. (2022). Stop With the Questions Already! Does Data Quality Suffer for Scales Positioned Near the End of a Lengthy Questionnaire? *Journal of Business and Psychology*, *37*, 1099–1116. https://doi.org/10.1007/s10869-021-09787-8

Bowling, N. A., Gibson, A. M., Houpt, J. W., & Brower, C. K. (2021). Will the Questions Ever End? Person-Level Increases in Careless Responding During Questionnaire Completion. *Organizational Research Methods*, *24*(4), 718–738. https://doi.org/10.1177/1094428120947794

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who Cares and Who Is Careless? Insufficient Effort Responding as a Reflection of Respondent Personality. *Journal of Personality and Social Psychology*, *111*(2), 218–229. http://dx.doi.org/10.1037/pspp0000085.supp

Bowling, N. A., Huang, J. L., Brower, C. K., & Bragg, C. B. (2023). The Quick and the Careless: The Construct Validity of Page Time as a Measure of Insufficient Effort Responding to Surveys. *Organizational Research Methods*, *26*(2), 323–352. https://doi.org/10.1177/10944281211056520

Braitman, A. L., Petrey, A. M., Shipley, J. L., Ayala Guzman, R., Renzoni, E., Looby, A., & Bravo, A. J. (2025). Check your data before you wreck your model: The impact of careless

responding on substance use data quality. *Alcohol, Clinical and Experimental Research*, *49*(4), 941–951. https://doi.org/10.1111/acer.70024

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). The Guilford Press.

Brower, C. K. (2018). *Too Long and Too Boring: The Effects of Survey Length and Too Long and Too Boring: The Effects of Survey Length and Interest on Careless Responding Interest* [Wright State University, Master's Thesis]. Core Scholar. https://corescholar.libraries.wright.edu/etd_all/1918

Brunet, J., Gunnell, K. E., Gaudreau, P., & Sabiston, C. M. (2015). An integrative analytical framework for understanding the effects of autonomous and controlled motivation. *Personality and Individual Differences*, *84*, 2–15. https://doi.org/10.1016/j.paid.2015.02.034

Brunet, J., Sabiston, C., Castonguay, A., Ferguson, L., & Bessette, N. (2012). The Association Between Physical Self-Discrepancies and Women's Physical Activity: The Mediating Role of Motivation. *Journal of Sport and Exercise Psychology*, *34*(1), 102–123. https://doi.org/10.1123/jsep.34.1.102

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*, *50*(6), 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for

cognition. *Psychological Bulletin*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor–Davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress*, *20*(6), 1019–1028. https://doi.org/10.1002/jts.20271

Camus, K. A. (2015). *Once Careless, Always Careless? Temporal and Situational Once Careless, Always Careless? Temporal and Situational Stability of Insufficient Effort Responding (IER) Stability of Insufficient Effort Responding (IER)* [Masters Thesis, Wright State University]. https://corescholar.libraries.wright.edu/etd_all

Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A New Scale for Measuring Adults' Prosocialness. *European Journal of Psychological Assessment*, *21*(2), 77–89. https://doi.org/10.1027/1015-5759.21.2.77

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Carver, C. S. (1997). You want to measure coping but your protocol's too long: Consider the brief COPE. *International Journal of Behavioral Medicine*, *4*(1), 92–100. https://doi.org/10.1207/s15327558ijbm0401_6

Cassidy, S. (2016). The Academic Resilience Scale (ARS-30): A New Multidimensional Construct Measure. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01787

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., Duriez, B., Lens, W., Matos, L., Mouratidis, A., Ryan, R. M., Sheldon, K. M., Soenens, B., Van Petegem, S., & Verstuyf, J. (2015). Basic psychological need satisfaction, need

frustration, and need strength across four cultures. *Motivation and Emotion*, *39*(2), 216–236. https://doi.org/10.1007/s11031-014-9450-1

Cheung, S. (2019). *Understanding Insufficient Effort Responding from a Self-Determination Theory Perspective* [Doctoral Thesis, University of Ottawa]. uO Research. https://ruor.uottawa.ca/items/5db7e905-2c10-42a6-a335-6bf73341c438

Cheval, B., Maltagliati, S., Courvoisier, D. S., Marcora, S., & Boisgontier, M. P. (2024). Development and validation of the physical effort scale (PES). *Psychology of Sport and Exercise*, *72*, 102607. https://doi.org/10.1016/j.psychsport.2024.102607

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385. https://doi.org/10.2307/2136404

Converse, J. M. (2009). *Survey Research in the United States* (1st ed.). Routledge. https://doi.org/10.4324/9781315130491

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, *3*(31), 2011–2079. https://doi.org/10.21500/20112084.844

Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise*, *35*(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Curran, P. G., & Hauser, K. A. (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality*, *82*, 103849. https://doi.org/10.1016/j.jrp.2019.103849

de Calvo, M. P. C., & Reich, D. A. (2007). Spontaneous correction in the behavioral confirmation process: The role of naturally-occurring variations in self-regulatory resources. *Basic and Applied Social Psychology*, *29*(4), 351–364. https://doi.org/10.1080/01973530701665132

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105–115. https://doi.org/10.1037/h0030644

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134. https://doi.org/10.1016/0092-6566(85)90023-6

DeSimone, J. A., Davison, H. K., Schoen, J. L., & Bing, M. N. (2018). Insufficient Effort Responding as a Partial Function of Implicit Aggression. *Organizational Research Methods*, *23*(1), 154–180. https://doi.org/10.1177/1094428118799486

DeSimone, J. A., & Harms, P. D. (2018). Dirty Data: The Effects of Screening Respondents Who Provide Low-Quality Data in Survey Research. *Journal of Business and Psychology*, *33*(5), 559–577. https://doi.org/10.1007/s10869-017-9514-9

Desimone, J. A., Harms, P. D., & Desimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171–181. https://doi.org/10.1002/job.1962

Diener, E., Emmons, R. A., Larsem, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences. *Journal of Business and Psychology*, *33*(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0

Dupuis, M., Meier, E., Capel, R., & Gendre, F. (2015). Measuring individuals' response quality in self-administered psychological tests: An introduction to Gendre's functional method. *Frontiers in Psychology*, *6*(May), 629. https://doi.org/10.3389/fpsyg.2015.00629

Dupuis, M., Meier, E., & Cuneo, F. (2018). Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices. *Behavior Research Methods*, *51*, 2228–2237. https://doi.org/10.3758/s13428-018-1103-y

Ekkekakis, P., Zenko, Z., & Vazou, S. (2021). Do you find exercise pleasant or unpleasant? The Affective Exercise Experiences (AFFEXX) questionnaire. *Psychology of Sport and Exercise*, *55*, 101930. https://doi.org/10.1016/j.psychsport.2021.101930

Elliot, A. J., & Thrash, T. M. (2010). Approach and Avoidance Temperament as Basic Dimensions of Personality. *Journal of Personality*, *78*(3), 865–906. https://doi.org/10.1111/j.1467-6494.2010.00636.x

Falk, C. F., Huang, A., & Ilagan, M. J. (2025). Unsupervised [randomly responding] survey bot detection: In search of high classification accuracy. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000746

Fischer, D. G., & Fick, C. (1993). Measuring Social Desirability: Short Forms of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, *53*(2), 417–424. https://doi.org/10.1177/0013164493053002011

Fraley, R. C., Heffernan, M. E., Vicary, A. M., & Brumbaugh, C. C. (2011). The experiences in close relationships—Relationship Structures Questionnaire: A method for assessing attachment orientations across relationships. *Psychological Assessment*, *23*(3), 615–625. https://doi.org/10.1037/a0022898

Francavilla, N. M., Meade, A. W., & Young, A. L. (2019). Social Interaction and Internet-Based Surveys: Examining the Effects of Virtual and In-Person Proctors on Careless Response. *Applied Psychology*, *68*(2), 223–249. https://doi.org/10.1111/apps.12159

Funke, F. (2016). A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales. *Social Science Computer Review*, *34*(2), 244–254. https://doi.org/10.1177/0894439315575477

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102. https://doi.org/10.1186/1479-5868-9-102

Gibson, A. M., & Bowling, N. A. (2020). The Effects of Questionnaire Length and Behavioral Consequences on Careless Responding. *European Journal of Psychological Assessment*, *36*(2), 410–420. https://doi.org/10.1027/1015-5759/a000526

Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *Leadership Quarterly*, *31*(4). https://doi.org/10.1016/J.LEAQUA.2020.101384

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology*, *48*(1), 82–98. https://doi.org/10.1037/0022-3514.48.1.82

Gratz, K. L., & Roemer, L. (2004). Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41–54. https://doi.org/10.1023/B:JOBA.0000007455.08539.94

Grau, I., Ebbeler, C., & Banse, R. (2019). Cultural Differences in Careless Responding. *Journal of Cross-Cultural Psychology*, *50*(3), 336–357. https://doi.org/10.1177/0022022119827379

Green-Demers, I., Pelletier, L. G., & Ménard, S. (1997). The impact of behavioural difficulty on the saliency of the association between self-determined motivation and environmental behaviours. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 29*(3), 157–166. https://doi.org/10.1037/0008-400X.29.3.157

Grimby, G., Börjesson, M., Jonsdottir, I. H., Schnohr, P., Thelle, D. S., & Saltin, B. (2015). The "Saltin–Grimby Physical Activity Level Scale" and its application to health research. *Scandinavian Journal of Medicine & Science in Sports*, *25*(S4), 119–125. https://doi.org/10.1111/sms.12611

Guertin, C., & Pelletier, L. (2023). Motivational dynamics in the prediction of self-regulation strategies for the quality and quantity of eating, eating behaviors, and life satisfaction: A longitudinal investigation. *Food Quality and Preference*, *107*, 104829. https://doi.org/10.1016/j.foodqual.2023.104829

Gummer, T., & Roßmann, J. (2014). Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review*, *33*(2), 217-234. https://doi.org/10.1177/0894439314533479

Harms, C., Jackel, L., & Montag, C. (2017). Reliability and completion speed in online questionnaires under consideration of personality. *Personality and Individual Differences*, *111*, 281–290. https://doi.org/10.1016/j.paid.2017.02.015

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2025). Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: Factors influencing careless responding. *Psychological Methods*, *30*(2), 374–392. https://doi.org/10.1037/met0000580

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, *21*(3), 360–373. https://doi.org/10.1177/0894439303253985

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Herzog, A. R., & Bachman, J. G. (1981). Effects of Questionnaire Length on Response Quality. *The Public Opinion Quarterly*, *45*(4), 549–559. https://doi.org/10.1086/268687

Hickman, C., Marks, E., Pihkala, P., Clayton, S., Lewandowski, R. E., Mayall, E. E., Wray, B., Mellor, C., & van Susteren, L. (2021). Climate anxiety in children and young people and their beliefs about government responses to climate change: a global survey. *The Lancet Planetary Health*, *5*(12), e863–e873. https://doi.org/10.1016/S2542-5196(21)00278-3

Hong, M., Steedle, J. T., & Cheng, Y. (2020). Methods of Detecting Insufficient Effort Responding: Comparisons and Practical Recommendations. *Educational and Psychological Measurement*, *80*(2), 312–345. https://doi.org/10.1177/0013164419865316

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and Deterring Insufficient Effort Responding to Surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

Ilagan, M. J., & Falk, C. F. (2023). Supervised Classes, Unsupervised Mixing Proportions: Detection of Bots in a Likert-Type Questionnaire. *Educational and Psychological Measurement*, *83*(2), 217–239. https://doi.org/10.1177/00131644221104220

Irish, K., & Saba, J. (2023). Bots are the new fraud: A post-hoc exploration of statistical methods to identify bot-generated responses in a corrupt data set. *Personality and Individual Differences*, *213*, 112289. https://doi.org/10.1016/j.paid.2023.112289

Jackson, D. N. (1977). *Jackson vocational interest survey: manual*. London, Ont.: Research Psychologists Press.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, *10*(JUN). https://doi.org/10.3389/fpsyg.2019.01258

Kam, C. C. S., & Meyer, J. P. (2015). How Careless Responding and Acquiescence Response Bias Can Influence Construct Dimensionality: The Case of Job Satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Karabatsos, G. (2003). Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics. *Applied Measurement in Education*, *16*(4), 277–298. https://doi.org/10.1207/S15324818AME1604_2

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

Kim, D., Lee, Y., Lee, J., Nam, J. K., & Chung, Y. (2014). Development of Korean Smartphone Addiction Proneness Scale for Youth. *PLoS ONE*, *9*(5), e97920. https://doi.org/10.1371/journal.pone.0097920

Kost, R. G., & Correa da Rosa, J. (2018). Impact of survey length and compensation on validity, reliability, and sample characteristics for Ultrashort-, Short-, and Long-Research Participant Perception Surveys. *Journal of Clinical and Translational Science*, *2*(1), 31–37. https://doi.org/10.1017/cts.2018.18

Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika*, *45*(2), 527–563. https://doi.org/10.1007/s41237-018-0063-y

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, *50*(1), 537–567. https://doi.org/10.1146/annurev.psych.50.1.537

Kwon, M., Kim, D.-J., Cho, H., & Yang, S. (2013). The Smartphone Addiction Scale: Development and Validation of a Short Version for Adolescents. *PLoS ONE*, *8*(12), e83558. https://doi.org/10.1371/journal.pone.0083558

Lavergne, K. J., & Pelletier, L. G. (2015). Predicting individual differences in the choice of strategy to compensate for attitude-behaviour inconsistencies in the environmental domain. *Journal of Environmental Psychology*, *44*, 135–148. https://doi.org/10.1016/j.jenvp.2015.10.001

Lavigne, G. L., & Vallerand, R. J. (2010). The Dynamic Processes of Influence Between Contextual and Situational Motivation: A Test of the Hierarchical Model in a Science Education Setting. *Journal of Applied Social Psychology*, *40*(9), 2343–2359. https://doi.org/10.1111/j.1559-1816.2010.00661

Lazarsfeld, P., & Fiske, M. (1938). The “Panel” as a New Tool for Measuring Opinion. *Public Opinion Quarterly*, *2*(4), 596. https://doi.org/10.1086/265234

Leiner, D. J. (2019). Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods*, *13*(3). https://doi.org/https://doi.org/10.18148/srm/2019.v13i3.7403

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December 15). Isolation forest. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 413–422. https://doi.org/10.1109/ICDM.2008.17

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, *2*, 49–55.

Malach-Pines, A. (2005). The Burnout Measure, Short Version. *International Journal of Stress Management*, *12*(1), 78–88. https://doi.org/10.1037/1072-5245.12.1.78

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*(1), 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, *84*(8), 79–83. https://doi.org/10.1016/j.paid.2014.08.021

Martin, A. J. (2013). Academic buoyancy and academic resilience: Exploring ‘everyday’ and ‘classic’ resilience in the face of academic adversity. *School Psychology International*, *34*(5), 488–500. https://doi.org/10.1177/0143034312472759

McCrae, R. R., & Costa, P. T. Jr. (2006). Cross-Cultural Perspectives on Adult Personality Trait Development. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 129–145). Lawrence Erlbaum Associates Publishers.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Meng, H., Cao, H., Hao, R., Zhou, N., Liang, Y., Wu, L., Jiang, L., Ma, R., Li, B., Deng, L., Lin, Z., Lin, X., & Zhang, J. (2020). Smartphone use motivation and problematic smartphone use in a national representative sample of Chinese adolescents: The mediating roles of smartphone use time for various activities. *Journal of Behavioral Addictions*, *9*(1), 163–174. https://doi.org/10.1556/2006.2020.00004

Miller, J. (1991). Short Report: Reaction Time Analysis with Outlier Exclusion: Bias Varies with Sample Size. *The Quarterly Journal of Experimental Psychology Section A*, *43*(4), 907–912. https://doi.org/10.1080/14640749108400962

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality*, *63*, 1–11. https://doi.org/10.1016/j.jrp.2016.04.010

Nimon, K., Zientek, L. R., & Henson, R. K. (2012). The Assumption of a Reliable Instrument and Other Pitfalls to Avoid When Considering the Reliability of Data. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00102

Olson, K., & Parkhurst, B. (2013). Collecting Paradata for Measurement Error Evaluations. In F. Kreuter (Ed.), *Improving Surveys with Paradata* (pp. 43–72). Wiley. https://doi.org/10.1002/9781118596869.ch3

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Ornstein, M. (2013). A Companion to Survey Research. In *A Companion to Survey Research* (pp. 1–10). SAGE Publications Ltd. https://doi.org/10.4135/9781473913943.n1

Pelletier, L. G., Dion, S. C., Slovinec-D'Angelo, M., & Reid, R. (2004). Why do you regulate what you eat? Relationships between forms of regulation, eating behaviors, sustained dietary behavior change, and psychological adjustment. *Motivation and Emotion*, *28*(3), 245–277. https://doi.org/10.1023/B:MOEM.0000040154.40922.14

Pelletier, L. G., Dion, S., Tuson, K., & Green-Demers, I. (1999). Why do people fail to adopt environmental protective behaviors? Toward a taxonomy of environmental amotivation. *Journal of Applied Social Psychology*, *29*(12), 2481–2504. https://doi.org/10.1111/j.1559-1816.1999.tb00122.x

Pelletier, L. G., Dion, S., Tuson, K., & Green-Demers, I. (1999). Why Do People Fail to Adopt Environmental Protective Behaviors? Toward a Taxonomy of Environmental Amotivation. *Journal of Applied Social Psychology*, *29*(12), 2481–2504. https://doi.org/10.1111/j.1559-1816.1999.tb00122.x

Pelletier, L. G., Legault, L. R., & Tuson, K. M. (1996). The Environmental Satisfaction Scale. *Environment and Behavior*, *28*(1), 5–26. https://doi.org/10.1177/0013916596281001

Pelletier, L. G., Tuson, K. M., Green-Demers, I., Noels, K., & Beaton, A. M. (1998). Why are you doing things for the environment? The Motivation Toward the Environment Scale (MTES). *Journal of Applied Social Psychology*, *28*(5), 437–468. https://doi.org/10.1111/j.1559-1816.1998.tb01714.x

Pelletier, L. G., & Rocchi, M. (2023). Organismic integration theory: A theory of regulatory styles, internalization, integration, and human functioning in society. In R. M. Ryan (Ed.), *The Oxford handbook of self-determination theory* (pp. 53–83). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197600047.013.4

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10 item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: A person-oriented analysis. *Journal of Educational Psychology*, *99*(4), 734–746. https://doi.org/10.1037/0022-0663.99.4.734

Rocchi, M. A., Wilson, P. M., Sylvester, B. D., Guerin, E., Shi, Z., & Sweet, S. N. (2023). Toward brief tools assessing motivation for exercise: Rationale and development of a six- and 12-item version of the Behavioral Regulation in Exercise 3 (BREQ3) Questionnaire. *Sport, Exercise, and Performance Psychology*, *12*(3), 205–227. https://doi.org/10.1037/spy0000321

Rocchi, M., Pelletier, L., Cheung, S., Baxter, D., & Beaudry, S. (2017). Assessing need-supportive and need-thwarting interpersonal behaviours: The Interpersonal Behaviours Questionnaire (IBQ). *Personality and Individual Differences*, *104*, 423–433. https://doi.org/10.1016/j.paid.2016.08.034

Rosenberg, M. (1965). Rosenberg Self-Esteem Scale. In *PsycTESTS Dataset*. https://doi.org/10.1037/t01038-000

Ryan, R. M. (1995). Psychological Needs and the Facilitation of Integrative Processes. *Journal of Personality*, *63*(3), 397–427. https://doi.org/10.1111/j.1467-6494.1995.tb00501.x

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037//0003-066x.55.1.68

Ryan, R. M., & Deci, E. L. (2006). Self-Regulation and the Problem of Human Autonomy: Does Psychology Need Choice, Self-Determination, and Will? *Journal of Personality*, *74*(6), 1557–1586. https://doi.org/10.1111/j.1467-6494.2006.00420.x

Ryan, R. M., & Deci, E. L. (2017). *Self-Determination Theory : Basic Psychological Needs in Motivation, Development, and Wellness*. The Guilford Press. https://doi.org/10.1521/978.14625/28806

Ryan, R. M., & Frederick, C. (1997). On Energy, Personality, and Health: Subjective Vitality as a Dynamic Reflection of Well-Being. *Journal of Personality*, *65*(3), 529–565. https://doi.org/10.1111/j.1467-6494.1997.tb00326.x

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, *45*, 736-750. https://doi.org/ 10.1037/0022-3514.45.4.736

Schimmack, U. (2020). A meta-psychological perspective on the decade of replication failures in social psychology. *Canadian Psychology*, *61*(4), 364–376. https://doi.org/10.1037/cap0000246

Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research*, *27*(4), 1077–1088. https://doi.org/10.1007/s11136-017-1767-2

Schroeders, U., Schmidt, C., & Gnambs, T. (2022). Detecting Careless Responding in Survey Data Using Stochastic Gradient Boosting. *Educational and Psychological Measurement*, *82*(1), 29–56. https://doi.org/10.1177/00131644211004708

Sheu, H.-B., Chong, S. S., Chen, H.-F., & Lin, W.-C. (2014). Well-being of Taiwanese and Singaporean college students: Cross-cultural validity of a modified social cognitive model. *Journal of Counseling Psychology*, *61*(3), 447–460. https://doi.org/10.1037/cou0000018

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, *15*(3), 194–200. https://doi.org/10.1080/10705500802222972

Soldatos, C. R., Dikeos, D. G., & Paparrigopoulos, T. J. (2000). Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. *Journal of Psychosomatic Research*, *48*(6), 555–560. https://doi.org/10.1016/S0022-3999(00)00095-7

Spence, G. B., & Oades, L. G. (2011). Coaching with self-determination in mind: Using theory to advance evidence-based coaching practice. *International Journal of Evidence Based Coaching and Mentoring*, *9*(2), 37–55.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder. *Archives of Internal Medicine*, *166*(10), 1092. https://doi.org/10.1001/archinte.166.10.1092

Spreng, R. N., McKinnon*, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to

Multiple Empathy Measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Standage, M., Duda, J. L., & Ntoumanis, N. (2003). A model of contextual motivation in physical education: Using constructs from self-determination and achievement goal theories to predict physical activity intentions. *Journal of Educational Psychology*, *95*(1), 97–110. https://doi.org/10.1037/0022-0663.95.1.97

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics, 5th ed. In *Using multivariate statistics, 5th ed.* Allyn & Bacon/Pearson Education.

Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations.* (H. Tajfel, Ed.). Academic Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37).

Tuckman, B. W. (1991). The Development and Concurrent Validity of the Procrastination Scale. *Educational and Psychological Measurement*, *51*(2), 473–480. https://doi.org/10.1177/0013164491512022

Ulitzsch, E., Pohl, S., Khorramdel, L., Kroehne, U., & von Davier, M. (2022). A Response-Time-Based Latent Response Mixture Model for Identifying and Modeling Careless and Insufficient Effort Responding in Survey Data. *Psychometrika*, *87*(2), 593–619. https://doi.org/10.1007/s11336-021-09817-7

United Nations Committee on the Rights of the Child. (2022). *General Comment No.26*. childrightsenvironment.org

Vallerand, R. J. (1997). Toward a Hierarchical Model of Intrinsic and Extrinsic Motivation. *Advances in Experimental Social Psychology*, *29*, 271–360. https://doi.org/10.1016/S0065-2601(08)60019-2

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C. B., & Vallières, E. F. (1992). Academic Motivation Scale. *Educational and Psychological Measurement*, *52*. https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft25718-000

Vansteenkiste, M., Lens, W., De Witte, S., De Witte, H., & Deci, E. L. (2004). The 'why' and 'why not' of job search behaviour: their relation to searching, unemployment experience, and well-being. *European Journal of Social Psychology*, *34*(3), 345–363. https://doi.org/10.1002/ejsp.202

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology*, *101*(3), 671–688. https://doi.org/10.1037/a0015083

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating Learning, Performance, and Persistence: The Synergistic Effects of Intrinsic Goal Contents and Autonomy-Supportive Contexts. *Journal of Personality and Social Psychology*, *87*(2), 246–260. https://doi.org/10.1037/0022-3514.87.2.246

Vansteenkiste, M., Zhou, M., Lens, W., & Soenens, B. (2005). Experiences of Autonomy and Control Among Chinese Learners: Vitalizing or Immobilizing? *Journal of Educational Psychology*, *97*(3), 468–483. https://doi.org/10.1037/0022-0663.97.3.468

Wang, M. D., & Hau, K.-T. (2024). Flagging insufficient effort responses in surveys: Stopping rule to prevent insufficient or excessive removal of doubtful data. *Acta Psychologica*, *243*, 104135. https://doi.org/10.1016/j.actpsy.2024.104135

Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417–430. https://doi.org/10.1016/J.CHB.2017.06.032

Ward, M. K., & Meade, A. W. (2023). Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices. *Annual Review of Psychology*, *74*(1). https://doi.org/10.1146/annurev-psych-040422-045007

Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417–430. https://doi.org/10.1016/J.CHB.2017.06.032

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender Differences in Personality across the Ten Aspects of the Big Five. *Frontiers in Psychology*, *2-2011*. https://doi.org/10.3389/fpsyg.2011.00178

Witt, E. A., Donnellan, M. B., & Orlando, M. J. (2011). Timing and selection effects within a psychology subject pool: Personality and sex matter. *Personality and Individual Differences*, *50*(3), 355–359. https://doi.org/10.1016/J.PAID.2010.10.019

Woods, C. M. (2006). Careless Responding to Reverse-Worded Items: Implications for Confirmatory Factor Analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3). https://doi.org/10.1007/s10862-005-9004-7

Yentes, R. D. (2020). *In Search of Best Practices for the Identification and Removal of Careless Responders* [Doctoral dissertation, North Carolina State University]. NC State Repository. http://www.lib.ncsu.edu/resolver/1840.20/37220

APPENDIX A

In order to ensure the replicability and transparency of our research, a detailed description of the measures and questionnaires used in the study can be found here. This documentation supplements the summary of the measures found in Table 3.2 in *Manuscript 2*.

Motivation Towards Research

Motivation for research participation was measured using the Motivation Towards Research scale (MTR; Cheung, 2019). The MTR is a 6-item measure which captures one's level of amotivation ("*I do not know why I am participating in research; it is a waste of my time*"), external ("*This is a required component of my course*"), introjected ("*I would feel guilty if I did not*"), identified ("*Participating in research is important*"), integrated ("*I personally value the importance of being a research participant*"), and intrinsic ("*Participating in research is interesting*") motivation towards research participation. Respondents were asked to rate each statement on a 7-point Likert scale, from 1 (*does not correspond at all*) to 7 (*corresponds entirely*), indicating how well each statement aligns with their reasons for participating in research. To obtain a composite score of autonomous motivation, the intrinsic, integrated, and identified items were averaged. To obtain a score of controlled motivation, the introjected and external items were averaged, while the amotivation item was interpreted as is. The internal consistency of the scale was found to be adequate in the present study ($\omega_t = .75$).

Global Motivation

Global motivation was measured using the 6-item version of the Global Motivation Scale (GMS; Pelletier et al., 2004). Each item from the GMS captures one's level of intrinsic ("*for the pleasure of acquiring new knowledge*"), integrated ("*because they reflect what I value the most in life*"), identified ("*because I choose them as means to attain my objectives*"), introjected

("because I would feel bad if I do not do them"), external ("*in order to show others what I am capable of*"), and amotivation ("*although it does not make a difference whether I do them or not*"), as defined by SDT. Participants are prompted to indicate the extent to which these statements represent the reason why they generally do things in their life (i.e., "*In general, I do things...*") using a 7-point Likert scale ranging from 1(*not agree at all*) to 7 (*completely agree*). To obtain a composite score of autonomous motivation, the intrinsic, integrated, and identified items were averaged. The introjected and external items were averaged to obtain a score of controlled motivation, while the amotivation item was interpreted as is. The internal consistency of the scale was found to be adequate in the present study ($\omega_t = .75$).

Personality

Participants' personality was assessed using the 10-item version of the Big Five Inventory (BFI; Rammstedt & John, 2007). Participants indicated their perceived level of agreeableness (e.g., "*is generally trusting*"), conscientiousness (e.g., "*does a thorough job*"), extraversion (e.g., "*is reserved*"), openness to experience (e.g., "*has an active imagination*") and neuroticism (e.g., "*gets nervous easily*") using a Likert scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Each of these traits was measured using two items. A composite score was calculated by firstly reversing negative items and then averaging the two items for each subscale. The internal consistency of the global scale was found to be adequate in the present study ($\omega_t = .74$). For the needs of the present study, solely the conscientiousness subscale was used.

Social Desirability

Participants social desirability was measured using the Marlowe-Crowne Social Desirability Scale - Short Form (Fischer & Fick, 1993). This scale contains 13 items (e.g., "*I never hesitate to go out of my way to help someone in trouble")* to which participants respond

using a True (0)/False (1) response scale. This scale was placed amongst other measures and was measured currently to IER. In order to obtain a measure of participants' level of social desirability, negative items (i.e., socially undesirable items) are reversed and a composite score is then obtained by summing all the items. Doing so returns a score contained between 0 and 13.

Attention Checks

In order to test the convergent validity of the mIERda, we analysed datasets containing three to five instructed items (e.g., "*Please respond 'Moderately Agree' to this item*") or bogus items (e.g., "I eat cement occasionally"; Huang et al., 2012; Meade & Craig, 2012). These items were placed evenly and inconspicuously across the survey by the researcher. After data collection, responses to these items were recoded as pass/fail. In the instance where a respondent provided the expected response their response was recoded as pass (0), otherwise the response was recoded as fail (1). To ensure that descriptive statistics were comparable across datasets, the proportion of failed attention checks was obtained by computing the mean of the recoded items; thus, yielding a score ranging between 0 and 1. These attention checks were otherwise pooled for analysis.

Self-Report Items

In order to further investigate the convergent validity of the mIERda, we analyzed datasets which contained responses to three self-report items (Meade & Craig, 2012). These items located at the end of the survey invited respondents to assess the level of effort (i.e., "*I put forth____ effort towards this study*"), attention (i.e., "*I gave this study ____ attention*") they provided, and lastly, whether they believed their data should be used by the research team (i.e., "*In your honest opinion, should we use your data? You will be fully compensated for your participation no matter your answer*"). Respondents responded to the effort and attention items

using a 5-point Likert scale ranging from 0 (*a lot of effort/attention*) to 4 (*almost no*). Respondents responded to self-reported data inclusion item using a 0 (*Yes*) to 1 (*No*) scale. Accordingly, a higher score on each item denotes a lower level of effort, attention and data quality.

Completion Time

For each dataset, the completion time in seconds was obtained during data collection using Qualtrics integrated timer. Given the extreme skewness (= 10.14) and kurtosis (= 131.87) of the distribution, a log transformation was applied to each dataset. Doing so improved the skewness (= 2.06) and kurtosis (= 9.85) of the overall distribution. To further improve the normality of the distribution, the data was winsorized at the 5th and 95th percentile of the completion time distribution for each dataset. Doing so reduced the skewness (= 1.08) and kurtosis (= 4.07) of the distribution.

Length of the survey

For each dataset, the length of the survey was obtained by computing the total number of items, irrespectively of the format, contained in the survey.

Time of the semester

To determine when a respondent completed the study in a given semester, the time of the semester was calculated by subtracting the date of participation from the date the subject pool opened (refer to Appendix C; Table C-3.2). Where the study was conducted, the subject pool opens seven days after the start of classes, so the opening date can be determined by adding seven days to the semester start date. The student pool opens annually on four occasions: September to December (fall semester; 12-13 weeks), January to April (winter semester; 12-13 weeks), May to June (spring semester; 5 weeks) and June to July (summer semester; 5 weeks).

Since the spring and summer semesters are largely shorter than the fall and winter semesters, the time of semester was treated as missing and subsequently imputed for datasets collected during those semesters. In order to ensure the comparability of time of semester across semesters and years, the variable was standardized.

APPENDIX B

Motivation Towards Research Scale (Cheung, 2019)

Please indicate to what extent each of the following statements correspond to the reasons why you participate in research / Veuillez indiquer dans quelle mesure chacune des déclarations suivantes correspond aux raisons pour lesquelles vous participez à la recherche.

1 Does not correspond at all/ Ne corresponds pas du tout	2 Corresponds very little/ Corresponds très peu	3 Corresponds a little/ Corresponds un peu	4 Corresponds moderately/ Corresponds modérément	5 Corresponds strongly/ Corresponds fortement	6 Corresponds very strongly/ Corresponds très fortement	7 Corresponds entirely/ Corresponds entièrement

Item	Rating
Participating in research is interesting/ Participer à des recherches est intéressant	1 2 3 4 5 6 7
Participating in research is important/ Participer à des recherches est important	1 2 3 4 5 6 7
I would feel guilty if I did not/ Je me sentirais coupable si je ne le faisais pas à	1 2 3 4 5 6 7
I personally value the importance of being a research participant/ J'accorde personnellement de l'importance au fait de participer à des recherches	1 2 3 4 5 6 7
This is a required component of my course/ C'est un élément obligatoire de mon cours	1 2 3 4 5 6 7
I do not know why I am participating in research; it is a waste of my time/ Je ne sais pas pourquoi je participe à des recherches; c'est une perte de mon temps	1 2 3 4 5 6 7

Global Motivation Scale (GMS-6; Cheung, 2019)

Indicate your level of agreement with each of the following statements, which corresponds generally to the reasons why you do different things./ Indiquez votre degré d'accord avec chacune des affirmations suivantes, qui correspondent de façon générale aux raisons pour lesquelles vous faites différentes choses.

In general, I do things.../ En général, je fais les choses...

1 Not agree at all/ Pas du tout en accord	2 Very slightly agree/ Très peu en accord	3 Slightly agree/ Un peu en accord	4 Moderately agree/ Moyennement en accord	5 Mostly agree/ Assez en accord	6 Strongly agree/ Très en accord	7 Completely agree/ Tout à fait en accord

Item							
...because they reflect what I value the most in life/ ...parce qu'elles reflètent ce que je valorise le plus dans la vie	1	2	3	4	5	6	7
…because I choose them as means to attain my objectives/ …parce que je les choisis comme moyens pour réaliser mes projets	1	2	3	4	5	6	7
...because I would feel bad if I do not do them/ …parce que je me sentirais mal de ne pas les faire	1	2	3	4	5	6	7
...for the pleasure of acquiring new knowledge/ …pour le plaisir d'acquérir de nouvelles connaissances	1	2	3	4	5	6	7
...in order to show others what I am capable of/ …pour montrer aux autres ce que je vaux	1	2	3	4	5	6	7
...although it does not make a difference whether I do them or not/ …bien que cela ne fasse pas de différence que je les fasse ou non	1	2	3	4	5	6	7

Big Five Inventory – 10 (BFI-10; Rammstedt & John, 2007)

How well do the following statements describe your personality?/ Dans quelle mesure les énoncés suivants décrivent-ils votre personnalité

1 Disagree Strongly/ Pas du tout en accord	2 Disagree a little/ Un peu en désaccord	3 Neither agree nor disagree/ Ni en accord ni en désaccord	4 Agree a little/ Un peu en accord	5 Agree strongly/ Tout à fait en accord

I see myself as someone who is reserved/ Je me vois comme quelqu'un qui est réservé	1	2	3	4	5
I see myself as someone who is generally trusting Je me vois comme quelqu'un qui fait généralement confiance aux autres	1	2	3	4	5
I see myself as someone who tends to be lazy/ Je me vois comme quelqu'un qui a tendance à être paresseux	1	2	3	4	5
I see myself as someone who is relaxed, handles stress well/ Je me vois comme quelqu'un qui est relaxe, détendu, gère bien le stress	1	2	3	4	5
I see myself as someone who has few artistic interests/ Je me vois comme quelqu'un qui est peu intéressé par tout ce qui est artistique	1	2	3	4	5
I see myself as someone who is outgoing, sociable/ Je me vois comme quelqu'un qui est sociable, extraverti	1	2	3	4	5
I see myself as someone who tends to find fault with others/ Je me vois comme quelqu'un qui a tendance à critiquer les autres	1	2	3	4	5
I see myself as someone who does a thorough job/ Je me vois comme quelqu'un qui travaille consciencieusement	1	2	3	4	5
I see myself as someone who gets nervous easily/ Je me vois comme quelqu'un qui est facilement anxieux	1	2	3	4	5
I see myself as someone who has an active imagination/ Je me vois comme quelqu'un qui a une imagination active	1	2	3	4	5

Self-Reported Measures of Insufficient Effort Responding (Meade & Craig, 2012)

Self-Reported Effort (Meade & Craig, 2012)

English Version

Lastly, it is vital to our study that we only include responses from people that devoted their full attention to this study. Otherwise, years of effort (the researchers' and the time of other participants) could be wasted. You will receive credit for this study no matter what, however, please tell us how much effort you put forth towards this study.

I put forth ____ effort towards this study
- almost no (4)
- very little (3)
- some (2)
- quite a lot of (1)
- a lot of (0)

Version Française

Enfin, il est essentiel pour notre étude que nous incluions seulement les réponses des personnes qui ont consacré toute leur attention à cette étude. Sinon, des années d'efforts (des chercheurs et le temps des autres participants) pourraient être perdues. Vous recevrez un crédit pour cette étude quoi qu'il arrive. Cependant, veuillez nous indiquer l'effort que vous avez consacré à cette étude.

J'ai fait ____ effort pour cette étude.
- pratiquement aucun (4)
- très peu d'(3)
- un peu d' (2)
- considérablement d'(1)
- beaucoup d'(0)

Self-Report Attention (Meade & Craig, 2012)

English Version

Also, often there are several distractions present during studies (other people, TV, music, etc.). Please indicate how much attention you paid to this study. Again, you will receive credit no matter what. We appreciate your honesty!

I gave this study ____ attention

- almost no (4)
- very little of (3)
- some of my (2)
- most of my (1)
- my full (0)

Version Française

De plus, il y a souvent plusieurs distractions présentes pendant les études (autres personnes, télévision, musique, etc.). Veuillez indiquer l'attention que vous avez portée à cette étude. Encore une fois, vous recevrez un crédit quoi qu'il arrive. Nous apprécions votre honnêteté !

J'ai prêté _______ attention à cette étude.

- pratiquement aucune (4)
- très peu d' (3)
- un peu d' (2)
- la plupart de mon (1)
- toute mon (0)

Self-Reported Data Inclusion (Meade & Craig, 2012)

English Version

In your honest opinion, should we use your data? You will be fully compensated for your participation no matter your answer.

- Yes (1)
- No (0)

Version Française

En toute honnêteté, pensez-vous que nous devrions utiliser vos réponses dans le cadre de notre recherche? Veuillez noter que, peu importe votre réponse, vous recevrez votre compensation pour votre participation à cette étude.

- Oui (1)
- Non (0)

APPENDIX C

Below, more information regarding the datasets used in *Manuscript 2* are provided. In Table C-3.1, we provide an overview of the objectives of each research program. Although the goal of *Manuscript 2* was not to assess these objectives or answer the research questions associated with each research program, this section aims to provide context and information to contextualize and enhance the understanding of the study's findings. In Table C-3.2, we provide an overview of the subject pool opening and closing date between Fall 2021 and Winter 2024 (i.e., data collection period). In Table C-3.3, we summarize the number of attention checks by datasets. The correct answer to each attention check is also specified. Tables C-3.4 to C-3.21 present the correlation between the mIERda status (start and end of the survey) and validation variables for each dataset. Lastly, Table C-3.22 presents the sensitivity of the mIERda with attention checks and self-reported items by research programs and datasets

Table C-3.1

Descriptive Information of Research Programs – Manuscript 2

Research Program	Main objective	Survey Length	Data collection period	*N* datasets
Academic Resilience	From this research, researchers sought to learn more about how students deal with difficult experiences during their university life. The goal was to get a better understanding of academic resilience, engagement, basic psychological needs, and motivation.	299	Fall 2022	4
Cellphone in the Classroom	From this research, researchers sought to understand how undergraduate and graduate students use their smartphones, for unrelated tasks, in the classroom.	164	Fall 2021	1
Motivation and Well-being	From this research, researchers sought to understand students' motivation and experiences in their daily life as they are completing their undergraduate studies	185-217	Spring 2022 to Winter 2023	4
Physical Activity	From this research, researchers sought to understand participants’ experiences with activities requiring physical effort and their relationship with motivation.	108-201	Fall 2022 to Winter 2024	5
Pro-Environmental Behaviours	From this research, researchers sought to enhance our comprehension of attitudes and behaviors toward ecology and climate changes. More specifically, this study sought for a clearer understanding of how factors related to close relationships, psychological well-being, and motivational system may be associated to current attitude toward ecology.	170-196	Fall 2023	4

Note. Survey length denotes the number of items contained within each dataset.

Table C-3.2

Subject Pool Opening and Closing Date by Semester

Semester	Year	Opening date	Closing date	*n* weeks	*n* datasets
Fall	2021	September 15th	December 8th	12 weeks	3
Winter	2022	January 17th	April 8th	11 weeks, 4 days	1
Spring	2022	May 9th	June 10th	4 weeks, 4 days	1
Summer	2022	June 27th	July 28th	4 week, 4 days	0
Fall	2022	September 14th	December 7th	12 weeks	5
Winter	2023	January 16th	April 12th	12 weeks, 2 days	0
Spring	2023	May 8th	June 9th	4 weeks, 4 days	1
Summer	2023	June 26th	July 28th	4 weeks, 4 days	1
Fall	2023	September 13th	December 6th	12 weeks	5
Winter	2024	January 15th	April 10th	12 weeks, 2 days	1

Note. The table below presents the opening and closing dates for each semester covered by the data collection period (Fall 2021-Winter 2024).

Table C-3.3

Summary of Attention Checks by Research Programs and Datasets

Variables	Cellphone in the Classroom	Physical Activity		Motivation & Well-being			Academic Resilience	Pro-Environmental Behaviours	
	Dataset 1	2	3 to 8	9	10 & 11	12	13 to 19	20 & 22	21 & 23
n Attention checks in survey	3	5	3	4	4	4	5	3	3
Attention check #1	1-2 range : 1-5 location: #71	7 range : 1-7 location: #56	2 range : 1-5 location: #44	2 range : 1-2 location: #146	3 range : 1-3 location: #61	2 range : 1-2 location: #61	1 range : 1-5 location: #29	1 range : 1-7 location: #56	1 range : 1-7 location: #29
Attention check #2	7 range : 1-7 location: #109	2 range : 1-2 location: #140	1-2 range : 1-7 location: #51	1 range : 1-7 location: #180	2 range : 1-2 location: #117	1 range : 1-7 location: #150	1 range : 1-2 location: #142	7 range : 1-7 location: #110	7 range : 1-7 location: #108
Attention check #3	1 range : 1-5 location: #143	1 range : 1-7 location: #164	1 range : 1-5 location: #69	1 range : 1-7 location: #192	1 range : 1-7 location: #151	1 range : 1-7 location: #164	3 range : 1-5 location: #181	4 range : 1-5 location #142	4 range : 1-5 location #152
Attention check #4		1 range : 1-7 location: #178		4 range : 1-7 location: #214	1 range : 1-7 location: #165	4 range : 1-7 location: #184	1 range : 1-7 location: #273		
Attention check #5		4 range : 1-7 location: #198					7 range : 1-7 location: #292		

Note. Values denote correct answer to attention check. Range refers to the Likert Scale of the item. Location refers to the position of the item (#) in the survey.

Table C-3.4

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 1

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.41* [.31, .50]	
Convergent Validity		
Attention checks	.55* [.53, .58]	.40* [.37, .43]
Self-reported effort	.06 [-.10, .21]	.10 [.-06, .26]
Self-reported attention	.32* [.24, .40]	.25* [.17, .34]
Self-reported data inclusion	.67* [.62, .72]	.50* [.43, .57]
Completion time[1,2]	-.24* [-.30, -.18]	-.11* [-.17, -.04]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	.02 [-.09, .04]	-.01 [-.07, -.06]
Research controlled motivation	.00 [-.03, .02]	-.02 [-.08, .04]
Predictive Validity		
Global autonomous motivation	-.12* [-.18, -.06]	.02 [-.05, .08]
Global amotivation	.12* [.06, .19]	.08* [.02, .15]
Research autonomous motivation	-.09* [-.16, -.03]	-.05 [-.11, .02]
Research amotivation	.17* [.11, .23]	.09* [.02, .15]
Conscientiousness	-.12* [-.18, -.06]	-.01 [-.07, .05]
Concurrent Validity		
Time of semester[1]	.11* [.04, .17]	.07 [.00, .13]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	-.01* [-.14, .09]	.09 [-.02, .21]

Note. n = 466–978. Results from the tetrachoric correlation (i.e., IER start-end, self-reported data inclusion, instructed items and gender) should be interpreted with caution due to the small sample size. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive; 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval.

[1]Standardized variable. [2]Log transformed variable.

Table C-3.5

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 2

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.60* [.47, .72]	
Convergent Validity		
Attention checks	.54* [.51, .57]	.48* [.45, .52]
Self-reported effort	.25* [.15, .33]	.22* [.13, .31]
Self-reported attention	.24* [.14, .33]	.22* [.13, .31]
Self-reported data inclusion	.52* [.45, .59]	.37* [.28, .45]
Completion time[1,2]	-.11 [-.20, -.01]	-.14* [-.23, -.04]
Discriminant Validity		
Social desirability	-.07 [-.15, .03]	.01 [-.09, .10]
Global controlled motivation	.02 [-.08, .11]	-.04 [-.13, .06]
Research controlled motivation	.00 [-.03, .02]	.02 [-.08, .11]
Predictive Validity		
Global autonomous motivation	-.06 [-.16, .03]	-.08 [-.18, .02]
Global amotivation	.23* [.13, .32]	.15* [.06, .25]
Research autonomous motivation	-.06 [-.16, .04]	-.01 [-.11, .09]
Research amotivation	.14* [.04, .23]	.08 [-.02, .18]
Conscientiousness	-.22* [-.31, -.12]	-.22* [-.31, -.12]
Concurrent Validity		
Time of semester[1]	.17* [.07, .26]	.06 [-.04, .15]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.17 [.02, .36]	.20 [.02, .37]

Note. n = 363–408. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive; 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable.

Table C-3.6

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 3

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.53* [.01, .84]	
Convergent Validity		
Attention checks	.62* [.49, .72]	.12 [-.06, .30]
Self-reported effort	.26 [-.05, .53]	.23 [-.09, .50]
Self-reported attention	.30 [-.01, .56]	.14 [-.18, .43]
Self-reported data inclusion	.62* [.38, .78]	.15 [-.17, .43]
Completion time[1,2]	-.07 [-.36, .24]	-.33 [-.57, -.02]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	.04 [-.27, .34]	-.07 [-.24, .36]
Research controlled motivation	.23 [-.08, .50]	.02 [-.08, .50]
Predictive Validity		
Global autonomous motivation	.01 [-.30, .31]	-.18 [-.46, .14]
Global amotivation	.10 [-.22, .40]	.13 [-.19, .43]
Research autonomous motivation	-.07 [-.37, .24]	.12 [-.19, .41]
Research amotivation	.10 [-.21, .40]	.08 [-.24, .37]
Conscientiousness	-.16 [-.45, .15]	-.08 [-.37, .23]
Concurrent Validity		
Time of semester	N/A	N/A
Survey length	N/A	N/A
Known-Groups Validity		
Gender	-.17 [-.64, .39]	-.22 [-.65, .29]

Note. n = 38–42. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive; 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable.

Table C-3.7

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 5

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.16 [-.52, .72]	
Convergent Validity		
Attention checks	.04 [-.19, .27]	.18 [-.05, .39]
Self-reported effort	.17 [-.21, .51]	.08 [-.31, .44]
Self-reported attention	.35 [-.03, .64]	.02 [-.36, .39]
Self-reported data inclusion	-.09 [-.45, .29]	.39 [.02, .67]
Completion time[1,2]	-.21 [-.54, .18]	-.25 [-.57, .13]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	.18 [-.21, .52]	.18 [-.21, .52]
Research controlled motivation	-.12 [-.47, .27]	.24 [-.15, .56]
Predictive Validity		
Global autonomous motivation	.16 [-.23, .50]	-.03 [-.40, .35]
Global amotivation	.24 [-.16, .56]	.32 [-.07, .62]
Research autonomous motivation	.63* [.33, .81]	.35 [-.03, .64]
Research amotivation	.03 [-.35, .40]	-.11 [-.47, .27]
Conscientiousness	-.14 [-.49, .24]	-.14 [-.49, .24]
Concurrent Validity		
Time of semester	N/A	N/A
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.38 [-.33, .83]	.38 [-.26, .82]

Note. n = 26–28. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.8

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 7

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.25 [.04, .44]	
	Convergent Validity	
Attention checks	.28* [.20, .35]	.07 [-.01, .15]
Self-reported effort	.11 [-.03, .24]	.08 [-.06, .21]
Self-reported attention	.15 [.01, .28]	.01 [-.13, .15]
Self-reported data inclusion	.40* [.28, .51]	-.12 [-.26, .02]
Completion time[1,2]	-.04 [-.17, .10]	.03 [-.11, .16]
	Discriminant Validity	
Social desirability	N/A	N/A
Global controlled motivation	-.10 [-.24, .04]	-.03 [-.17, .10]
Research controlled motivation	.06 [-.08, .20]	-.09 [-.22, .05]
	Predictive Validity	
Global autonomous motivation	-.06 [-.20, .07]	.01 [-.13, .14]
Global amotivation	.00 [-.14, .14]	.01 [-.13, .14]
Research autonomous motivation	.05 [-.09, .18]	-.07 [-.21, .07]
Research amotivation	-.05 [-.19, .09]	-.01 [-.15, .13]
Conscientiousness	-.11 [-.24, .03]	-.19* [-.32, -.05]
	Concurrent Validity	
Time of semester[1]	-.05 [-.19, .09]	-.02 [-.16, .12]
Survey length	N/A	N/A
	Known-Groups Validity	
Gender	.00 [-.26, .29]	-.05 [-.31, .21]

Note. n = 257–293. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.9

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 8

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.54* [.36, .67]	
Convergent Validity		
Attention checks	.33* [.27, .39]	.35* [.29, .40]
Self-reported effort	.15* [.04, .26]	.12 [.01, .23]
Self-reported attention	.23* [.12, .34]	.25* [.13, .35]
Self-reported data inclusion	.43* [.33, .52]	.41* [.31, .50]
Completion time[1,2]	-.06 [-.17, .06]	-.03 [-.14, .09]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	.03 [-.09, .15]	-.03 [-.08, .15]
Research controlled motivation	.01 [-.10, .13]	-.02 [-.13, .10]
Predictive Validity		
Global autonomous motivation	-.03 [-.15, .09]	-.04 [-.15, .08]
Global amotivation	.01 [-.10, .13]	.03 [-.09, .14]
Research autonomous motivation	-.07 [-.18, .05]	-.13 [-.24, -.02]
Research amotivation	.01 [-.10, .13]	.04 [-.08, .15]
Conscientiousness	-.15 [-.26, -.03]	-.22* [-.32, -.10]
Concurrent Validity		
Time of semester[1]	.07 [-.04, .18]	-.01 [-.12, .11]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.13 [-.11, .34]	-.05 [-.26, .17]

Note. n = 180–202. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.10

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 9

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.44* [.21, .63]	
Convergent Validity		
Attention checks	.45* [.39, .51]	.45* [.39, .50]
Self-reported effort	.11 [-.03, .25]	.12 [-.03, .25]
Self-reported attention	.18* [.04, .32]	.26* [.12, .38]
Self-reported data inclusion	.50* [.38, .59]	.63* [.53, .71]
Completion time[1,2]	-.19* [-.32, -.05]	-.13 [-.27, .00]
Discriminant Validity		
Social desirability	-.01 [-.14, .12]	-.05 [-.20, .09]
Global controlled motivation	.03 [-.11, .17]	-.02 [-.16, .12]
Research controlled motivation	-.02 [-.15, .12]	-.04 [-.18, .10]
Predictive Validity		
Global autonomous motivation	-.11 [-.25, .03]	-.04 [-.18, .10]
Global amotivation	.09 [-.05, .23]	.16 [.02, .29]
Research autonomous motivation	-.13 [-.26, .01]	-.08 [-.22, .06]
Research amotivation	.13 [-.01, .26]	.07 [-.07, .21]
Conscientiousness	-.11 [-.25, .03]	-.10 [-.24, .04]
Concurrent Validity		
Time of semester	N/A	N/A
Survey length	N/A	N/A
Known-Groups Validity		
Gender	-.21 [-.51, .11]	.15 [-.21, .39]

Note. n = 149–197. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.11

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 10

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.48* [.34, .64]	
Convergent Validity		
Attention checks	.32* [.26, .37]	.22* [.17, .28]
Self-reported effort	N/A	N/A
Self-reported attention	N/A	N/A
Self-reported data inclusion	N/A	N/A
Completion time[1,2]	.00 [-.12, .12]	-.01 [-.13, .11]
Discriminant Validity		
Social desirability	-.10 [-.22, .03]	-.08 [-.19, .03]
Global controlled motivation	-.02 [-.14, .10]	.02 [-.11, .14]
Research controlled motivation	-.00 [-.13, .12]	.03 [-.09, .15]
Predictive Validity		
Global autonomous motivation	-.20* [-.31, -.08]	-.15 [-.26, -.03]
Global amotivation	.15 [.03, .27]	.19* [.07, .30]
Research autonomous motivation	-.10 [-.22, .02]	-.12 [-.24, .00]
Research amotivation	.00 [-.12, .12]	.09 [-.04, .21]
Conscientiousness	-.21* [-.33, -.10]	-.16* [-.28, -.05]
Concurrent Validity		
Time of semester[1]	.10 [-.02, .22]	.10 [-.02, .22]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.10 [-.15, .35]	.01 [-.22, .25]

Note. n = 243–268. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.12

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 11

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.71* [.51, .85]	
Convergent Validity		
Attention checks	.66* [.61, .71]	.66* [.61, .71]
Self-reported effort	.13 [-.05, .31]	.22* [.03, .39]
Self-reported attention	.26* [.08, .43]	.30* [.12, .46]
Self-reported data inclusion	.22* [.05, .37]	.44* [.29, .56]
Completion time[1,2]	-.04 [-.20, .13]	-.09 [-.26, .08]
Discriminant Validity		
Social desirability	.00 [-.18, .21]	.09 [-.06, -.24]
Global controlled motivation	-.05 [-.21, .12]	.10 [-.06, .27]
Research controlled motivation	.11 [-.06, .27]	.04 [-.13, .20]
Predictive Validity		
Global autonomous motivation	-.07 [-.23, .10]	-.04 [-.20, .10]
Global amotivation	.13 [-.03, .30]	.21 [.05, .37]
Research autonomous motivation	-.10 [-.27, .07]	.01 [-.16, .17]
Research amotivation	.03 [-.14, .20]	-.04 [-.21, .13]
Conscientiousness	-.05 [-.22, .11]	-.31* [-.45, -.15]
Concurrent Validity		
Time of semester[1]	.20 [.03, .35]	.28* [.12, .43]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.43* [.12, .07]	.06 [-.31, .38]

Note. n = 109–137. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.13

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 12

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.59* [.49, .69]	
Convergent Validity		
Attention checks	.66* [.63, .68]	.65* [.63, .68]
Self-reported effort	.28* [.20, .36]	.17* [.08, .26]
Self-reported attention	.24* [.15, .33]	.25* [.16, .33]
Self-reported data inclusion	.59* [.53, .65]	.59* [.53, .65]
Completion time[1,2]	-.13* [-.21, -.04]	-.16* [-.25, -.07]
Discriminant Validity		
Social desirability	-.08 [-.18, .01]	-.10 [.21, -.02]
Global controlled motivation	-.04 [-.13, .05]	.03 [-.06, .12]
Research controlled motivation	-.11 [-.19, -.01]	-.08 [-.17, .01]
Predictive Validity		
Global autonomous motivation	-.18* [-.27, -.09]	-.09 [-.18, .00]
Global amotivation	.12 [.03, .21]	.19* [.10, .27]
Research autonomous motivation	-.06 [-.15, .03]	-.07 [-.16, .02]
Research amotivation	.18* [.09, .27]	.23* [.14, .31]
Conscientiousness	-.26* [-.34, -.17]	-.24* [-.33, -.15]
Concurrent Validity		
Time of semester[1]	.10 [.00, .18]	.12 [.03, .21]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.12 [-.06 .31]	.03 [-.12, .21]

Note. n = 420–468. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.14

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 13

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.30* [.08, .52]	
Convergent Validity		
Attention checks	.10* [.03, .16]	.49* [.44, .54]
Self-reported effort	.11 [-.02, .24]	.20* [.07, .33]
Self-reported attention	.07 [-.06, .21]	.18* [.04, .30]
Self-reported data inclusion	.11 [-.02, .24]	.43* [.31, .53]
Completion time[1,2]	.04 [-.10, .19]	-.12 [-.25, .01]
Discriminant Validity		
Social desirability	-.16 [-.27, .03]	-.14 [-.27, .01]
Global controlled motivation	-.11 [-.24, .03]	.09 [-.04, .23]
Research controlled motivation	-.02 [-.16, .11]	.06 [-.07, .20]
Predictive Validity		
Global autonomous motivation	-.22* [-.34, -.08]	-.07 [-.20, .07]
Global amotivation	.09 [-.04, .23]	-.01 [-.14, .12]
Research autonomous motivation	-.04 [-.17, .10]	-.09 [-.22, .04]
Research amotivation	.11 [-.02, .24]	.19 [.06, .32]
Conscientiousness	-.22* [-.34, -.09]	-.08 [-.21, .05]
Concurrent Validity		
Time of semester[1]	-.02 [-.15, .12]	.00 [-.13, .14]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.08 [-.21 .31]	.01 [-.27, .26]

Note. n = 188–214. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.15

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 17

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.58 [-.07, .90]	
Convergent Validity		
Attention checks	.71* [.60, .79]	.62* [.49, .72]
Self-reported effort	.22 [-.16, .55]	.44 [.08, .70]
Self-reported attention	.36 [-.01, .65]	.62* [.32, .81]
Self-reported data inclusion	.84* [.69, .93]	.91* [.82, .96]
Completion time[1,2]	.06 [-.32, .41]	-.30 [-.60, .08]
Discriminant Validity		
Social desirability	.04 [-.37, .51]	-.04 [-.40, .34]
Global controlled motivation	.08 [-.30, .43]	.12 [-.26, .46]
Research controlled motivation	.08 [-.29, .44]	.15 [-.23, .49]
Predictive Validity		
Global autonomous motivation	-.04 [-.40, .33]	-.24 [-.56, .14]
Global amotivation	.05 [-.33, .40]	.25 [-.13, .56]
Research autonomous motivation	.01 [-.36, .38]	-.18 [-.52, .19]
Research amotivation	-.20 [-.53, .18]	.27 [-.10, .58]
Conscientiousness	-.25 [-.57, .13]	-.34 [-.63, .03]
Concurrent Validity		
Time of semester[1]	-.25 [-.57, .13]	.15 [-.23, .49]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.11 [-.55 .58]	.02 [-.53, .52]

Note. n = 28–29. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.16

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 18

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.58 [.08, .85]	
Convergent Validity		
Attention checks	.82* [.77, .86]	.62* [.54, .70]
Self-reported effort	.39* [.14, .60]	.15 [-.12, .40]
Self-reported attention	.37* [.11, .59]	.17 [-.10, .42]
Self-reported data inclusion	.58* [.37, .73]	.55* [.33, .71]
Completion time[1,2]	-.47* [-.66, -.24]	-.14 [-.40, .13]
Discriminant Validity		
Social desirability	-.12 [-.32, .10]	-.13 [-.38, .10]
Global controlled motivation	-.06 [-.33, .21]	-.08 [-.34, .19]
Research controlled motivation	.13 [-.38, .15]	-.01 [-.28, .26]
Predictive Validity		
Global autonomous motivation	-.02 [-.29, .25]	-.13 [-.38, .10]
Global amotivation	.02 [-.25, .29]	-.16 [-.12, .41]
Research autonomous motivation	.01 [-.26, .27]	-.13 [-.39, .14]
Research amotivation	.17 [-.11, .42]	.13 [-.15, .38]
Conscientiousness	-.17 [-.42, .10]	-.23 [-.47, .04]
Concurrent Validity		
Time of semester[1]	.00 [-.26, .27]	.00 [-.27, .27]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.16 [-.45, .66]	-.10 [-.54, .33]

Note. n = 48–54. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.17

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 19

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.66* [.43, .83]	
Convergent Validity		
Attention checks	.60* [.54, .66]	.59* [.53, .65]
Self-reported effort	.23 [.04, .40]	.30 [.12, .47]
Self-reported attention	.23 [.05, .41]	.20 [.01, .38]
Self-reported data inclusion	.51* [.35, .64]	.74* [.65, .82]
Completion time[1,2]	-.30* [-.46, -.12]	-.12 [-.30, .08]
Discriminant Validity		
Social desirability	-.10 [-.28, .08]	-.11 [.-29, .08]
Global controlled motivation	.05 [-.14, .24]	.02 [-.17, .21]
Research controlled motivation	-.08 [-.26, .12]	-.05 [-.23, .14]
Predictive Validity		
Global autonomous motivation	-.16 [-.34, .03]	-.07 [-.25, .12]
Global amotivation	.12 [-.07, .31]	.24 [.05, .41]
Research autonomous motivation	-.10 [-.28, .09]	-.15 [-.33, .04]
Research amotivation	.22 [.03, .40]	.29* [.11, .45]
Conscientiousness	-.27* [-.44, -.09]	-.04 [-.23, .15]
Concurrent Validity		
Time of semester[1]	.03 [-.16, .22]	.04 [-.15, .23]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.01 [-.32, .33]	.05 [-.30, .41]

Note. n = 94–107. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.18

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 20

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.38* [.27, .50]	
Convergent Validity		
Attention checks	.59* [.55, .62]	.27* [.23, .32]
Self-reported effort	.19* [.11, .27]	.14* [.06, .22]
Self-reported attention	.19* [.11, .27]	.13* [.05, .21]
Self-reported data inclusion	.41* [.44, .48]	.21* [.13, .28]
Completion time[1,2]	-.14* [-.22, -.06]	-.06 [-.15, .02]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	.05 [-.03, .13]	.01 [-.08, .09]
Research controlled motivation	.06 [-.02, .14]	.01 [-.07, .09]
Predictive Validity		
Global autonomous motivation	-.06 [-.15, .03]	.01 [-.07, .10]
Global amotivation	.10* [.02, .19]	.01 [-.08, .09]
Research autonomous motivation	.04 [-.04, .12]	-.12* [-.20, -.04]
Research amotivation	.05 [-.03, .13]	.07 [-.01, .15]
Conscientiousness	-.13* [-.21, -.04]	-.14* [-.22, -.06]
Concurrent Validity		
Time of semester[1]	.11* [.03, .19]	.02 [-.06, .10]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.03 [-.12, .18]	-.15 [-.29, .01]

Note. n = 549–598. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.19

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 21

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.43* [.29, .54]	
Convergent Validity		
Attention checks	.41* [.37, .45]	.48* [.44, .51]
Self-reported effort	.10 [.02, .18]	.22* [.14, .29]
Self-reported attention	.14* [.06, .22]	.22* [.14, .29]
Self-reported data inclusion	.27* [.20, .34]	.63* [.58, .68]
Completion time[1,2]	-.14* [-.22, -.07]	-.25* [-.32, -.17]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	-.05 [-.13, .03]	.04 [-.05, .12]
Research controlled motivation	-.06 [-.14, .02]	-.03 [-.12, .05]
Predictive Validity		
Global autonomous motivation	-.03 [-.11, .05]	.04 [-.04, .12]
Global amotivation	.07 [-.01, .15]	.12* [.04, .20]
Research autonomous motivation	-.00 [-.08, .08]	.04 [-.04, .12]
Research amotivation	.05 [-.03, .13]	.11 [.03, .19]
Conscientiousness	-.12* [-.20, -.04]	-.07* [-.14, .02]
Concurrent Validity		
Time of semester[1]	.11 [-.03, .19]	.11 [.02, .18]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.00 [-.14, .15]	.12 [.00, .24]

Note. n = 515–567. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.20

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 22

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.36* [.16, .55]	
Convergent Validity		
Attention checks	.75* [.71, .79]	.28* [.21, .36]
Self-reported effort	.38* [.26, .49]	.21* [.08, .34]
Self-reported attention	.39* [.26, .49]	.17 [.04, .30]
Self-reported data inclusion	.56* [.46, .65]	.16 [.03, .29]
Completion time[1,2]	-.30* [-.42, -.17]	-.10* [-.24, .03]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	.03 [-.10, .17]	.02 [-.11, .16]
Research controlled motivation	-.04 [-.17, .10]	.04 [-.09, .18]
Predictive Validity		
Global autonomous motivation	-.20* [-.32, -.06]	-.13 [-.26, .01]
Global amotivation	.20* [.07, .33]	.04 [-.10, .17]
Research autonomous motivation	-.09 [-.22, .05]	-.06 [-.20, .07]
Research amotivation	.22* [.09, .35]	.20* [.07, .33]
Conscientiousness	-.17 [-.30, -.03]	-.13 [-.26, .00]
Concurrent Validity		
Time of semester[1]	.19* [.05, .31]	.12 [-.02, .25]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.46* [.27, .64]	.23 [.03, .43]

Note. n = 184–210. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable.

Table C-3.21

Correlation between the multidimensional Insufficient Effort Responding detection approach and Validation Variables – Dataset 23

Variables	IER	
	Start of the Survey	End of the Survey
IER (Start - End)	.46* [.26, .61]	
Convergent Validity		
Attention checks	.57* [.51, .62]	.43* [.37, .50]
Self-reported effort	.29* [.16, .41]	.36* [.23, .47]
Self-reported attention	.32* [.19, .44]	.38* [.26, .49]
Self-reported data inclusion	.54* [.43, .63]	.63* [.54, .71]
Completion time[1,2]	-.23* [-.36, -.10]	-.20* [-.33, -.07]
Discriminant Validity		
Social desirability	N/A	N/A
Global controlled motivation	-.01 [-.15, .12]	.02 [-.12, .15]
Research controlled motivation	.05 [-.09, .18]	-.03 [-.16, .11]
Predictive Validity		
Global autonomous motivation	-.10 [-.23, .03]	-.10 [-.23, .04]
Global amotivation	.05 [-.09, .18]	.08 [-.05, .22]
Research autonomous motivation	-.04 [-.17, .10]	-.03 [-.17, .10]
Research amotivation	.16* [.03, .29]	.11 [-.03, .24]
Conscientiousness	-.10 [-.23, .03]	-.06 [-.19, .07]
Concurrent Validity		
Time of semester[1]	.11 [-.03, .24]	.06 [-.08, .19]
Survey length	N/A	N/A
Known-Groups Validity		
Gender	.08 [-.17, .33]	.26 [.03, .49]

Note. n = 192–214. Results should be interpreted with caution due to the small sample size and lack of statistical power. * $p < .01$. The correlations were estimated using the IER status as estimated by the mIERda (0 = Attentive, 1 = IER). Self-reported data inclusion (0 = include, 1 = remove). Gender (0 = women, 1 = men). The confidence interval denotes the 95% confidence interval. [1]Standardized variable. [2]Log transformed variable. N/A: Variable unavailable

Table C-3.22

mIERda Overlap/Sensitivity with Attention Checks and Self-Reported Items by Research Programs and datasets

Variables	Cellphone in the Classroom	Physical Activity					Motivation & Well-being				Academic Resilience				Pro-Environmental Behaviours			
	Dataset 1	2	3	5	7	8	9	10	11	12	13	17	18	19	20	21	22	23
Number of failed attention checks																		
≥ 1 Attention Check	62.2	53.5	66.6	60.0	57.9	37.2	67.9	64.0	50.0	80.0	66.7	44.0	40.0	44.4	53.3	60.7	66.6	45.7
≥ 2 Attention Checks	88.0	77.8	75.0	75.0	81.0	47.5	37.5	54.5	75.0	75.0	66.7	100	75.0	63.6	78.2	80.0	82.6	64.1
≥ 3 Attention Checks	91.7	81.2	100	–	100	76.1	91.7	100	100	100	87.5	100	100	78.6	93.4	97.8	97.1	100
≥ 4 Attention Checks		100					100	100	100	100	100	100	100	100				
5 Attention Checks		100									–	–	–	100				
Self-Reported Data Inclusion																		
"Don't use my data"	90.9	77.3	85.7	66.7	76.9	67.5	87.5	N/A	58.8	89.2	66.7	100	70.0	90.9	77.8	87.0	85.0	87.5

Note. Values denote the sensitivity of the mIERda (i.e., proportion of participants flagged by the indicator that were also captured by the mIERda) in %. – : Denotes instances where no participants failed X number of attention checks; N/A: Measure not included in the dataset.

APPENDIX D

In Table D-1, more information regarding the measures contained within the datasets analyzed in *Manuscript 2* and *Manuscript 3* are provided. It is worth noting that no participant appears in both manuscripts. Even if the data arises from the same datasets, each dataset was divided into two distinct samples. Participants who only completed one study were included in the sample analyzed in *Manuscript 2*. Participants who completed two or more studies were included in the longitudinal sample analyzed in *Manuscript 3*.

Table D-1

Summary of Measures by Research Programs and Datasets – Manuscript 2 & 3

Measures	Research Program																							
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours				
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23	
Academic Motivation Scale (Vallerand et al., 1992), 28 items, 7 subscales, 1-7 scale	X	X							X	X	X	X												
Academic Risk and Resilience Scale (Martin, 2013), 10 items, 0-1 scale													X	X	X	X	X	X	X					
Academic Resilience Scale (Cassidy, 2016), 30 items, 3 subscales, 1-5 scale													X	X	X	X	X	X	X					
Academic Stress Scale, (Sheu et al., 2014), 10 items, 0-4 scale	X																							
Affective Exercise Experiences (Ekkekak is et al., 2021), 6			X	X	X	X	X	X																

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
items, 2 subscales, 1-7 scale																							
Amotivation Towards the Environmental Scale (Pelletier, Dion, Tuson, & Green-Demers, 1999), 13 items, 4 subscales, 1-7 scale																				X		X	
Approach-Avoidance Temperament Questionnaire (Elliot & Thrash, 2010), 12 items, 2 subscales, 1-7 scale			X	X	X	X	X	X															
Athens Insomia Scale (Soldatos et al., 2000), 8 items, 1-4 scale	X																						
Basic Psychological Needs Satisfaction and Frustration Scale (Chen et al., 2015), 24	X	X							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
items, 6 subscales, 1-5 scale																							
Basic Psychological Needs Satisfaction and Frustration Scale (Chen et al., 2015), 6 items, 6 subscales, 1-5 scale		X							X	X	X	X											
Behavioral Regulation in Exercise 3, (M. A. Rocchi et al., 2023), 12 items, 6 subscales, 0-4 scale		X	X	X	X	X	X	X															
Brief Resilience Scale (Smith et al., 2008), 6 items, 1-7 scale																					X		X
Burnout Measure - Short Version (Malach-Pines, 2005), 10 items, 1-7 scale		X																					
Center for Epidemiologic Studies																				X		X	

	Research Program																						
Measures	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
Depression Scale - 10-item version (Andresen et al., 1994), 10 items, 1-4 scale																							
Child Friendly Environmental Questionnaire (United Nations Committee on the Rights of the Child, 2022), 3 items, All that Apply scale																				X	X	X	X
Connor Davidson Resilience Scale (Campbell-Sills & Stein, 2007), 10 items, 0-4 scale													X	X	X	X	X	X	X				
Coping Orientation to Problems Experienced Inventory (Brief-COPE; Carver, 1997), 28 items, 14 subscales, 0-3 scale													X	X	X	X	X	X	X				

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
Difficulties in Emotion Regulation Scale - Short Version (Gratz & Roemer, 2004), 18 items, 1-4 scale																				X		X	
Environmental Satisfaction Scale (Pelletier et al., 1996), 8 items, 2 subscales, 1-7 scale																					X		X
Experiences in Close Relationships, (Fraley et al., 2011), 9 items, 2 subscales, 1-7 scale																				X	X	X	X
Frequency of Pro-Environmental Behaviours (Lavergne & Pelletier, 2015), 9 items, open-ended scale																				X	X	X	X

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
GAD-7 (Spitzer et al., 2006), 7 items, 0-3 scale	X												X	X	X	X	X	X	X	X	X	X	X
General Causality Orientations Scale, (Deci & Ryan, 1985), 36 items, 3 subscales, 1-7 scale									X														
Global Motivation Scale (Pelletier et al., 2004), 18 items, 6 subscales, 1-7 scale									X	X	X	X											
Global Motivation Scale (Pelletier et al., 2004), 6 items, 6 subscales, 1-7 scale		X	X	X	X	X	X	X	X	X	X	X											
How Do You Feel Toward the Environment (Pelletier et al., 1996), 12 items, 2 subscales, 1-7 scale																				X	X	X	X

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
Inconsistency Compensation Strategies Scale (Lavergne & Pelletier, 2015), 9 items, 3 subscales, 1-7 scale																					X		X
Intolerance to Uncertainty Scale - Short Version (Carleton et al., 2007), 12 items, 2 subscales, 1-5 scale																					X		X
International Physical Activity Questionnaire (Craig et al., 2003), 4 items, open-ended scale			X	X	X	X	X	X													X		X
Interpersonal Behaviour Scale (M. Rocchi et al., 2017), 24 items, 6 subscales, 1-7 scale		X							X	X	X	X	X	X	X	X	X	X	X				

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
Interpersonal Behaviour Scale – Short Form (Rocchi et al., 2017), 6 items, 6 subscales, 1-7 scale		X							X	X	X	X											
Marlow-Crowne Social Desirability Scale, (Andrews & Meyer, 2003), 10 items, 0-1 scale		X							X	X	X	X	X	X	X	X	X	X	X				
Motivation Towards the Environment Scale (Pelletier et al., 1998), Short Form, 6 items, 6 subscales, 1-7																				X	X	X	X
Perceived Competence for Environmental Behaviours (Pelletier et al., 1998), 6 items, 1-7 scale																					X		X
Perceived Stress Scale (S. Cohen et al.,	X																			X		X	

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
1983), 10 items, 1-5 scale																							
Physical Activity Level Scale (Grimby et al., 2015), 1 item, 1-4 scale		X	X	X	X	X	X	X															
Physical Effort Scale (Cheval et al., 2024), 18 items, 2 subscales, 1-5 scale		X	X	X	X	X	X	X															
Positive Affects and Negative Affects Scale (Watson et al., 1988), 11 items, 2 subscales, 1-7 scale		X							X	X	X	X											
Procrastination Scale (Tuckman, 1991), 35 items, 2 subscales, 1-4 scale													X	X	X	X	X	X	X				
Pro-Environmental Attitudes (Lavergne & Pelletier, 2015), 8																				X	X	X	X

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
items, 2 subscales, 1-7 scale																							
Prosocialness Scale for Adults (Caprara et al., 2005), 16 items, 2 subscales, 1-5 scale																				X		X	
Need for Cognition (Cacioppo et al., 1996), 6 items, 1-5 scale			X	X	X	X	X																
Relationship Questionnaire (Bartholomew & Horowitz, 1991), 8 items, 4 subscales, 1-4,1-7, 1-8 scales																				X	X	X	X
Rosenberg Self-Esteem Scale (Rosenberg, 1965), 13 items, 0-1 scale		X							X	X	X	X											
Satisfaction with Life Scale (Diener et al.,		X							X	X	X	X											

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
1985), 5 items, 1-7 scale																							
Self-Efficacy Scale (Ajzen, 1991), 3 items, 1-5 scale			X	X	X	X	X	X															
Self-Report Habit Index - Automaticity subscale (Gardner et al., 2012), 4 items, 1-7 scale			X	X	X	X	X	X															
Smartphone Addiction Scale (Kwon et al., 2013), 10 items, 1-6 scale	X																						
Smartphone Addiction Proneness Scale (Kim et al., 2014), 15 items, 1-4 scale	X																						
Smartphone Use Motivation (H. Meng et al., 2020), 13 items, 3 subscales, 1-5 scale	X																						

Measures	Research Program																						
	Cellphone in the Classroom	Physical Activity							Motivation & Well-being				Academic Resilience							Pro-Environmental Behaviours			
	Dataset 1	2	3	4*	5	6*	7	8	9	10	11	12	13	14*	15*	16*	17	18	19	20	21	22	23
Subjective Vitality Scale (Ryan & Frederick, 1997), 9 items, 1-7 scale		X							X	X	X	X											
Toronto Empathy Questionnaire (Spreng et al., 2009), 16 items, 1-5 scale																				X		X	

Note. *Solely analyzed in *Manuscript 3*. Due to other miscellaneous items, socio-demographic questions, and since some measures were administered more than once, the sum of items might not equal the total survey length

APPENDIX E

Below, more information regarding the datasets analyzed in *Manuscript 3* are provided. In Table E-4.1, we provide an overview of the objectives of each research program. Although the goal of the present study was not to assess these objectives or answer the research questions associated with each research program, this section aims to provide context and information to contextualize and enhance the understanding of our study’s findings. In Table E-4.2, I provide an overview of the subject pool opening and closing date between Fall 2021 and Winter 2024 (i.e., data collection period). In Table E-4.3, I summarize the socio-demographic information by datasets. Lastly, Table E-4.4 presents the prevalence of IER, at the start and the end of the survey, for each research program.

Table E-4.1

Descriptive Information of Research Programs – Manuscript 3

Research Program	Main objective	Survey Length	Data collection period	*N* datasets
Academic Resilience	From this research, researchers sought to learn more about how students deal with difficult experiences during their university life. The goal was to get a better understanding of academic resilience, engagement, basic psychological needs, and motivation.	299	Fall 2022	7
Cellphone in the Classroom	From this research, researchers sought to understand how undergraduate and graduate students use their smartphones, for unrelated tasks, in the classroom.	164	Fall 2021	1
Motivation and Well-being	From this research, researchers sought to understand students' motivation and experiences in their daily life as they are completing their undergraduate studies	185-217	Spring 2022 to Winter 2023	4
Physical Activity	From this research, researchers sought to understand participants' experiences with activities requiring physical effort and their relationship with motivation.	108-201	Fall 2022 to Winter 2024	7
Pro-Environmental Behaviours	From this research, researchers sought to enhance our comprehension of attitudes and behaviors toward ecology and climate changes. More specifically, this study sought for a clearer understanding of how factors related to close relationships, psychological well-being, and motivational system may be associated to current attitude toward ecology.	170-196	Fall 2023	4

Note. Survey length denotes the number of items contained within each dataset.

Table E-4.2

Subject Pool Opening and Closing Date by Semester

Semester	Year	Opening date	Closing date	*n* weeks	*n* datasets
Fall	2021	September 15th	December 8th	12 weeks	3
Winter	2022	January 17th	April 8th	11 weeks, 4 days	1
Spring	2022	May 9th	June 10th	4 weeks, 4 days	1
Summer	2022	June 27th	July 28th	4 week, 4 days	0
Fall	2022	September 14th	December 7th	12 weeks	8
Winter	2023	January 16th	April 12th	12 weeks, 2 days	0
Spring	2023	May 8th	June 9th	4 weeks, 4 days	2
Summer	2023	June 26th	July 28th	4 weeks, 4 days	2
Fall	2023	September 13th	December 6th	12 weeks	5
Winter	2024	January 15th	April 10th	12 weeks, 2 days	1

Note. The table below presents the opening and closing dates for each semester covered by the data collection period (Fall 2021-Winter 2024).

Table E-4.3

Socio-Demographic Information by Research Programs – Manuscript 3

	Research Program 1		Research Program 2		Research Program 3		Research Program 4		Research Program 5	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender										
Women	474	65.4	247	65.9	291	63.5	1002	65.4	619	66.8
Men	163	22.5	72	19.2	89	19.4	374	24.4	214	23.1
Other	43	5.9	25	6.7	27	5.9	76	5.0	37	4.0
Age										
16-19 years	516	71.2	265	70.7	324	70.6	1025	66.7	673	72.7
20-25 years	178	24.6	87	23.2	114	24.8	482	31.6	253	27.3
26 years and older	20	2.8	12	3.2	12	2.4	10	0.7	0	0.0
Ethnicity										
BIPOC	392	54.1	213	56.8	253	55.2	836	54.6	475	51.3
White	310	42.8	149	39.7	188	41.1	651	42.5	424	45.8
Total										
Participants	725	100	377	100	458	100	1532	100	926	100

Note. Descriptive statistics for gender age and ethnicity denote the number of distinct participants, not of observations. The proportion might not sum to 100% due to missingness. Research program 1 = Academic Resilience, Research program 2 = Cellphone use in Classroom, Research program 3 = Motivation and Well-being, Research program 4 = Physical Activity, Research program 5 = Pro-environmental behaviours. BIPOC = Black, Indigenous and People of Colour

Table E-4.4

Frequency and prevalence of IER as per the Multidimensional Insufficient Effort Responding Approach by Research Programs – Manuscript 3

Research Program	Start of survey		End of survey	
	n	%	*n*	%
Research program 1	363	26.2	421	30.4
Research program 2	119	31.6	139	36.7
Research program 3	176	35.6	170	34.3
Research program 4	712	38.7	813	44.2
Research program 5	309	33.1	335	35.9

Note. IER classification was established using the mIERda (0 = Attentive, 1 = IER). Research program 1 = Academic Resilience, Research program 2 = Cellphone use in Classroom, Research program 3 = Motivation and Well-being, Research program 4 = Physical Activity, Research program 5 = Pro-environmental behaviours.

APPENDIX F

Ethics Approval

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

CERTIFICAT D'APPROBATION ÉTHIQUE I CERTIFICATE OF ETHICS APPROVAL

Numéro du dossier / Ethics File Number	H-04-25-11034
Titre du projet / Project Title	Garbage In, Garbage Out – Measuring and Predicting Insufficient Effort Responding in Survey Research
Type de projet / Project Type	Thèse de doctorat / Doctoral thesis
Statut du projet / Project Status	Approuvé / Approved
Date d'approbation (jj/mm/aaaa) / Approval Date (dd/mm/yyyy)	02/04/2025
Date d'expiration (jj/mm/aaaa) / Expiry Date (dd/mm/yyyy)	01/04/2026

Équipe de recherche / Research Team

Chercheur / Researcher	Affiliation	Role
Ariane GAUTHIER	École de psychologie / School of Psychology	Chercheur Principal / Principal Investigator
Luc PELLETIER	École de psychologie / School of Psychology	Superviseur / Supervisor

Conditions spéciales ou commentaires / Special conditions or comments

550, rue Cumberland, pièce 154
Ottawa (Ontario) K1N 6N5 Canada

550 Cumberland Street, Room 154
Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie I www.recherche.uottawa.ca/ethics