Advancing Statistics Education Among Social Sciences Students: Investigating the Relations between Statistical and Spatial Cognition as well as Existing Tests on Statistics Understanding

Rose-Marie Gibeau

Thesis submitted to the University of Ottawa

In partial fulfillment of the requirements for the

Ph. D. in experimental psychology

School of Psychology

Faculty of Social Sciences

University of Ottawa

Acknowledgments

As my doctorate comes to an end, excitement and apprehension boils in my stomach. I am impatiently waiting for what comes next and I know that my family is too. They think I did six years of research alone in a dark university office, but what they don't know is that I succeeded partly because of their support and their help. My partner, Marie-Pier Pelletier, was my rock throughout this academic journey. She never stopped supporting me through my growth and my recurring discouragement. It is because of our many conversations that this thesis is the way it is now. My sister, Marianne Chabot, was especially patient with me during my bachelor's degree, where I "death-stared" her when she dared interrupting me in the middle of a study session. My mother, Carole Louise Chabot, is the reason why I could embark such a journey. I inherited her curiosity and creativity, which are essential in research. As these traits have manifested as a passion for art in her life, they have manifested as a reluctant search for answers in mine.

I would like to thank my supervisor, Denis Cousineau, who took a chance on a young woman (me) who almost failed her statistics courses and welcomed her in his lab for her Honours' thesis. I am forever grateful for him to see my potential and to invest time and (lab) money in my education and growth. It is his advice and mentoring that allowed me to develop my research skills and my confidence as a researcher.

My supervisor was not the only one who helped me grow as a researcher. There was Erin Maloney, Lucile Chanquoy and Sébastien Béland who were great collaborators and even better mentors. I am grateful to have had such an elite research team who taught me how to be a good researcher and who also showed me what scientific integrity looks like. There was also Martin Lalumière whose comments on my thesis made it better and allowed me to squeeze some more learning into my brain, which made me a better producer and consumer of statistics. Thank you to Isabelle Boutet, Martin Lalumière, Patrick Gaudreau and Mustafa Baloğlu who graciously accepted to be members of my evaluation committee. It is an honor to have you on this last step of my journey.

Lastly, I would like to thank all the other collaborators and my senior lab-mates as it really does take a village to train a good researcher.

General Abstract

The goal of this thesis is to improve the education of statistics. The main idea of this contribution comes from *Numerical Cognition*, which relates mathematics to space. Since mathematics is related to statistics, could there be a study of *Statistical Cognition* examining the role of space in statistics education? It is the goal that this thesis will examine through three studies. Study A examines the relations between spatial, mathematics and statistics anxieties. A total of 778 participants – recruited from Canadian, French and Belgian universities – were asked to complete a questionnaire measuring their spatial, mathematics and statistics anxieties. The results indicate that all variables are related, and that spatial anxiety predicts statistics anxiety. Study B looks at visuospatial working memory (Study 1; $N = 306$) and spatial problem-solving skills (Study 2; $N = 374$), and their relation to statistics anxiety. Using the same population, we asked participants to complete the same questionnaire as well as two tasks (one measuring visuospatial working memory and one measuring verbal working memory). In the second study, we asked participants to complete the spatial and statistics anxiety measures as well as two other tasks (one measuring verbal abilities and one measuring spatial problem-solving skills) The results indicate that spatial skills are related to and predict interpretation anxiety in statistics. In order to examine the possibility of statistical cognition, one needs an instrument measuring statistics understanding. In order to develop such a test, one needs to know which statistical concepts it will evaluate. As such, Study C examines this prerequisite. A total of 42 statistics instructors were recruited to tell us which concepts were being measured in each item of four tests measuring statistics understanding. The analysis of their responses resulted in a list of 18 concepts. These studies indicate that spatial skills could be related to the understanding of statistics, which contributes to the foundation for a statistical cognition field of research.

Table of Contents

Chapter 1: General Introduction

Defining Statistics

When asked "what are statistics?" different people will have different answers. There is no single good answer as to what statistics is. Edgeworth (1885) defined statistics as three things: "the arithmetical portion of Social Science" (p. 181), "the science of Means in general (including physical observations)" (p. 182) and the science of "Means which are presented by social phenomena" (p. 182). The author further argued that this *Science of Means* is divided in two based on the type of problem it is solving. One of these two types of problems looks at mean difference and whether this difference occurred by chance or not. The second type of problems looks at the best kind of mean (e.g., arithmetical, geometrical) that should be used. Of course, current statistics involves a lot more than the simple *Science of Means*.

Then, there is Willcox (1936), who listed 117 definitions in different languages from 1749 to 1934. The author ended the article with his own definition: "statistics has a broader meaning, the numerical study of groups, and a narrower one, the numerical study of social groups" (p. 399). His definition summarizes a common theme across the listed definitions, i.e., the science of facts that occur within a society or a group of people, which is similar to Edgeworth (1885).

The above definitions were very pertinent in their days and rooted in the origin of statistics. Statistics was first used by governments to factually describe their affairs (Hayslett, 1968). Today, statistics is not only used by governments. As such, defining statistics as the *Science of Social Facts* may not be as relevant now. Or maybe it is as relevant; the definition just needs to be more inclusive of statistical techniques and information that are now used. Specifically, it is no longer a science of means or of social facts only.

Hayslett (1968) argued that statistics is a field of mathematics and that they "are quantities that have been calculated from sample data” (p. 5). Some researchers would argue against this first part. Moore and Cobb (2000), for example, explicitly said that statistics use mathematics, but they are not a field of this last subject. The authors said that there are many reasons why that is. One of them is that context is essential in statistics, and a source of confusion in mathematics.

More recently, Rice et al. (2024), defined statistics in a way that reflects this separation from mathematics while using its equations. They argued that statistics are “the application of statistical tests, and procedures for organizing, analyzing and interpreting results or data according to mathematical formulas” (p. 210). The focus of this definition is put on what is being done and how it is done, which is rooted in mathematics. However, with today’s technology, statistics no longer have to be rooted in mathematics. Indeed, mathematical formulas and *t* or *z* tables are not necessarily consulted or used nowadays, especially by hand (Carver et al., 2016). Indeed, technologies (e.g., statistical software) are increasingly used, however inconsistently as there are many different types all with their particular advantages and disadvantages (Zavez & Harel, 2025).

One communality between these different definitions across eras is that statistics are the science of quantities based on social groups or society that are calculated by tests and procedures. Within this *Science of Quantitative Social Facts*, there are people who do the science, people who consume the facts, people who teach this science and people who learn this science. The present thesis will focus not on the last two groups of people (the ones who are teaching or learning), but on the act of teaching and learning the subject.

Statistics Education

Teaching statistics should not be taken lightly for both the teachers’ benefits and their students'. In fact, even after formal training, many students still hold misconceptions. Indeed, researchers and non-researchers alike tend to think in simplistic ways (Tversky & Kahneman, 1971). It is then essential to teach statistics effectively and to train students to be aware of their misconceptions. As will be discussed in Chapters 2 to 4 of the present thesis, both the teaching and learning of statistics are challenging for many reasons. These reasons involve the curriculum, learning goals, assessment and teaching methods, and students' simplistic ways of reasoning. The preliminary version of the subsections below is available in PsyArXiv (https://osf.io/preprints/psyarxiv/arv3q_v1).[1]

Curricula

In mathematics education, there are building blocks. Children are being gradually taught the necessary concepts to allow a deeper understanding of this subject (Cousineau & Harding, 2017). However, researchers have not yet determined those building blocks in statistics, which is one of the reasons why the teaching is difficult. A consequence of this is curricula reforms. Many researchers have called for reforms of statistical curricula. For example, Moore (1997) suggested that statistics courses need to be more balanced between "data analysis, data production, and inference" (p.126).

The examination of statistical curricula can be divided into four categories. First, the focus of the curriculum, the order of the curriculum, the topics included in the curriculum, and, lastly, the topics missing in the curriculum.

The Focus of the Curriculum is Misplaced

This wave of reform has generated many questions related to the concepts that educators seek to teach (e.g., Falk & Konold, 1993). There is not yet consensus regarding those concepts in

[1] Gibeau, R.-M., & Cousineau, C. (2021). Statistics Education Reboot : Starting Point. *PsyArXiv*. Retrieved from https://osf.io/preprints/psyarxiv/arv3q_v1

an introductory statistics course. Some researchers proposed to teach “the big picture” in order to develop students’ statistical reasoning skills (Zieffler et al., 2008, p.11).

In a study, Mustafa (1996) developed a statistics course curriculum for an undergraduate business program. While going through the curriculum, it is possible to take out the key concepts that are being taught. First, the question concerning the utility and nature of statistics is addressed. Then, there are the actual topics taught in this order: Data collection; psychometry; descriptive statistics; normal distribution; population, sample and sample distribution; statistics distributions and errors; confidence intervals and margin of error; hypothesis testing; graphs; controlled experiments; randomisation and experimental design; modeling; linear model; regression model; the limits of models.

This outline seems to be a mix of research methods and statistics. Interestingly, graphs are taught at the end of the class. Students learn to compute, quantify and dichotomize the data first and learn to visualize the data last. According to the Guidelines for Assessment and Instruction in Statistics Education II (GAISE II; Bargagliotti et al., 2020), students should be presented with data visualization methods earlier (see their recommendations for Level A on pp. 22-23). Also, the only statistical school of thought that is covered is Null Hypothesis Significance Testing (NHST). Other statistical schools, like Bayesian statistics for example, are omitted. However, note that the *p*-value is not the only statistic used to measure the probability of an estimate taught in the outline. Indeed, a measure of precision, confidence intervals, is also included.

The Curriculum is not Properly Ordered

Other authors proposed resequencing the curricula (Malone et al., 2010). This alternative way of teaching operates on four principles: “more closely mimic what a scientist/statistician does" (p.53), "get to statistical inference earlier" (p.54), "follow what we know from learning theory" (p.54) and "teach just enough probability to get by" (p.54).

The idea behind Principle 1 is that when a statistician is a consultant on a project, their work is not linear (e.g., graphs for all research questions, hypothesis tests for all research questions). Instead, they might do graphs at different occasions. The second principle is based on Wardrop's (1995) textbook, where hypothesis testing is located as early as Chapter 2. Principle 3 encourages instructors to use what we learned from research on learning theories. Specifically, the authors said that the concepts should be repeated in different contexts. In Principle 4, the authors encouraged instructors to only teach basic probability concepts which will support the understanding of *p*-values.

The Curriculum is Full of Unnecessary Topics

The Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report (Carver et al., 2016) added a section about concepts that should not be taught in statistics courses. For examples, creating graphs by hand is no longer necessary with today's technology and the same goes for teaching *z* or *t* tables. The GAISE II also recommended that technology be an integral part of statistics education (Bargagliotti et al., 2020). The use of technology was promoted much earlier in Coulter (1995). Instead of manual calculation, time should be devoted to developing a deeper understanding of graph and *p*-value interpretation (rather than teaching how to compute a statistic by hand). Considering the controversy surrounding the *p*-value, it may be interesting to teach the meaning of it as well as its issues and practical solutions.

One can wonder why the same problems (e.g., technology use, the controversy of the *p*-value, etc.) are still current more than 50 years later despite progress in teaching methods. According to Bargagliotti et al. (2020), this is because the discipline of statistics is constantly evolving. Traditional teaching approaches are still used and are increasingly being questioned. Tomkins and Ulus (2016), however, proposed to use Experiential Learning when teaching statistics. This proposition concerns how to teach statistical concepts instead of what concepts to teach. Perhaps it is useful to include in statistics education as this approach specifically focus on how students can “make their own sense of the content, and craft their own connections among the various concepts” (p.159). Supporting this, other researchers argued that introductory statistics courses should not focus on specific statistical methods (Billard, 1998; Curry, 2002; Moore, 1997). Similarly to Tomkins and Ulus’s (2016) recommendation, a recent study examined the benefits of gamification in statistics education (Khoshnoodifar et al., 2023). This active learning method is “the use of game design elements” (p.231) in teaching. Those elements can include “narrative, avatar, level, points, progress bars, scoreboards, challenges, and feedback” (p.233).

The Curriculum is Missing Necessary Topics

It is also possible that something is missing in statistics curricula, something that would help students learn and understand the connections between concepts. For Meletiou (2000), what is missing is an emphasis on variability. She argued that because variability is often ignored, students do not learn to recognize and understand uncertainty and variation. In Bargagliotti et al.’s (2020) recommendations, variability and its display methods are also at the center of data analysis for Level B students (see III. Analyze Data on p. 45).

However, adding topics might be difficult as there are some constraints involved in statistics teaching that needs to be considered. For example, Harth (2018) mentioned that time constraints must be considered while developing a course – most introductory statistics courses are one-semester long. In fact, she considered four constraints when it comes to statistics education. Namely, “the institutional, didactic, pedagogic and curricular constraints” (p.101). The institutional constraints concern the concepts that are covered or strategies that are used. The didactic constraints revolve around student engagement. Pedagogical constraints are at the intersection of the student learning style and the professor pedagogical style and philosophy. Finally, the amount of time that can be given to statistics is limited. This means that not all statistical concepts can be covered in a single course.

Expectations and Assessment

Learning Goals

Just as there is no consensus on the concepts to teach, there is also no agreement regarding what students should learn or what specific skills they should acquire by the end of a course. However, there seems to be an agreement regarding general learning requirements. One can look at them as being the purpose of statistics courses. These are statistical thinking, statistical literacy and statistical reasoning (Broers, 2006). However, the definition of each term is not definitive. Many researchers have different ways of defining and operationalizing those skills.

Statistical thinking can be considered as “a fundamental method of enquiry that cannot simply be learned but that has to mature in talented individuals who by nature possess this type of intelligence” (Broers, 2006, p. 3). This definition needs to be taken carefully as it is elitist, which can give way to discrimination. Statistical literacy is often defined as the understanding of basic statistical language or terminology and of statistical concepts. Finally, statistical reasoning can be understood as how people understand statistics; how they reason when facing statistical concepts or problems. This involves the understanding of basic concepts (e.g., spread, uncertainty, etc.) and interpreting results.

Although those broad objectives are often talked about (e.g., Broers, 2006), Gal and Garfield (1997) suggested different learning goals. Those are, 1) “understand the purpose and logic of statistical investigations”; 2) “understand the process of statistical investigations”; 3) “master procedural skills”; 4) “understand mathematical relationships”; 5) “understand probability and chance”; 6) “develop interpretative skills and statistical literacy”; 7) “develop ability to communicate statistically”; 8) “develop useful statistical disposition” (pp. 3-5).

The first learning goal concerns the understanding of the purpose of statistics (why it is used). The second concerns the understanding of the statistics process and of the scientific method as well as when certain statistical processes should be used. The third goal involves the ability to compute statistics, build graphs and tables, and organize data. The fourth goal concerns the understanding of the underlying mathematics of statistical concepts and tests. Fifth, this goal involves understanding just enough probability to understand statistical inference. The sixth goal concerns students’ ability to interpret statistical results as well as their limitations and possible biases. The seventh goal concerns students’ writing and speaking abilities to communicate statistical results. Lastly, students should understand chance and randomness as well as their role in statistics.

On the other hand, Stuart (1995) argued that students must be able to solve statistical problems. As such, each step of the statistical problem-solving process must be taught to students, and not only data analysis. Specifically, problem formulation, statistical design, data collection, interpretation and implementation must be included. According to the author, problem formation gives the basis from which students can build their understanding of statistics. It does so by leading to questions requiring data analysis. In other words, students should be led by curiosity and appreciation of statistics (Billard, 1998; Curry, 2002).

More recently, Berndt et al. (2021), suggested that two essential learning goals are statistical literacy to critically evaluate statistical information, and scientific reasoning and argumentation skills to use scientific concepts and methods in different contexts. The former is domain-specific, and the latter is general (i.e., science-focused). These learning goals are necessary for evidence-based decision making. However, the authors argued that they are underdeveloped. Some influencing factors would be numeracy skills, and the beliefs or ideas about knowledge and knowing.

Assessing Students

Factors to Consider. Assessment is the biggest and the most feared part of a statistics courses (Onwuegbuzie & Seaman, 1995). To improve the learning of statistics, researchers might want to look at what influences test performance. Those factors are, for example, time constraints during examinations, self-efficacy, students' motivation, lexical ambiguity (Finney & Schraw, 2003; Holmes, 2014; Jung & Hwang, 2016; Onwuegbuzie & Seaman, 1995). Statistics anxiety is also a factor to consider in statistics education as it can influence students' understanding (Cruise et al., 1985; Lawton 1994). As will be seen in Study A of Chapter 2 of the present thesis, the best predictor, however, seems to be statistics anxiety (Cantinotti et al., 2017). As will be seen in Study B of Chapter 3 of the present thesis, some studies also indicate that working memory influences performance (e.g., Almarzouki, 2024; Bergman Nutley & Söderqvist, 2017).

It is essential that assessments properly evaluate students' comprehension considering that assessments initiate and give momentum to students for learning the topics (Hubbard, 1997). According to the author, instructors "should assess what they have been asking their students to learn and in the way they envisage the students have learned it" (p. 2). As such, if an instructor taught the mean and assumes students learn the concept with formulas, they should not evaluate students' learning of the mean with graphs. Also, it has been proposed to use different types of questions in evaluations. Using only one type encourages students to memorize statistical methods instead of engaging in deeper learning processes.

Another factor instructors must be aware of and try to minimize is statistical test anxiety, which is different from general test anxiety (Benson, 1989). One way to reduce anxiety, proposed by Onwuegbuzie and Seaman (1995), is to make sure students have enough time to complete the test. A further problem related to assessment in statistics education is students' motivation. Harth (2018) proposed to mention the learning objectives for each lesson and on each assessments, which can be based on the subject (i.e., relates to the content of the lessons), on personal transferable skills (i.e., communicating numeracy and working in teams) and on generic skills (i.e., relates to critical thinking, analyzing and summarizing hypotheses).

How to Build Assessments. DelMas (2002) suggested assessing students with certain key words related to one of the three general learning requirements (i.e., statistical thinking, statistical literacy and statistical reasoning). To assess students' statistical thinking the terms "*apply*, *critique*, *evaluate*, *generalize*" should be used; for statistical literacy one can use the terms "*identify*, *describe*, *re-phrase*, *translate*, *interpret*, *read*; finally, statistical reasoning can be assessed with the terms "*why*, *how*, *explain* (a process)" (Table 1 of pp. 6-7). This suggestion is informative, but it does not provide concrete ways to evaluate students and how to minimize the factors which affect performance.

Garfield and Ben-Zvi (2008) provided concrete ways to evaluate students depending on the instructor's purpose for the assessment. This purpose then indicates which assessment method(s) is(are) most appropriate between the following eight: Homework, quizzes and exams, projects, critiques and writing assessments, lab reports, collaborative assessment tasks, minute papers, and student reflections and dispositions. Another author examined the benefits of three teaching strategies on students' performance (Marson, 2007). Specifically, he looked at repetition, immediate feedback and the use of original data. He concluded that all three positively impact students' learning, but immediate feedback is the most beneficial as it allows students to see their errors and compare them to the correct answers. As such, assessment methods should be used thoughtfully and should allow rapid feedback.

Even though no study, to my knowledge, have examined university policies on evaluations, it is safe to say that many courses end with a final exam. In statistics education, most tests evaluate students' ability to compute numbers (Allen, 2006). As will be seen in Chapter 5 of the present thesis, considering that statistics is more than the mere memorization of formulas, conceptual understanding may be more essential to evaluate than computational and procedural skills. The debate on formulas and evaluating computational skills is not new. Pollatsek et al. (1981) argued that learning formulas is a poor strategy to develop an understanding of underlying concepts. Additionally, the GAISE College Report recommended to stop assessing students' computational skills, especially by hand (Carver et al., 2016). Bargagliotti et al. (2020) also recommended to assess conceptual understanding of students no matter the format of evaluations used.

As will be seen in Study C of Chapter 4 of the present thesis, in an attempt to find measures of conceptual understanding, a thorough search has been carried out. Only four tests were found that specifically measure conceptual understanding of college or university students. Specifically, there is the *Statistical Reasoning Assessment* (SRA; Garfield, 2003), the *Statistics Concept Inventory* (SCI; Allen, 2006), the *Comprehensive Assessment of Outcomes for a first* *course in Statistics* (CAOS; DelMas et al., 2007) and the *Statistical Reasoning in Biology Concept Inventory* (SRBCI; Deane et al., 2016). When looking at the items (97 in total), it is not clear what they measure exactly. Some measure conceptual understanding of concepts, but others still evaluate formulas and memorized definitions.

Students' Simplistic Ways of Reasoning

Known Heuristics

Kahneman and Tversky (1972) were the firsts to examine the incorrect simplistic ways humans reason about uncertainty. Even though people do have some understanding and knowledge of basic probability and randomness, heuristics are used to evaluate complex real-world problems or events in a less energy- and time-consuming way (Konold, 1989). In an article, three heuristics are described (Tversky & Kahneman, 1974).

The first heuristic described is the *representativeness heuristic*. This way of reasoning concerns the probability of an event to occur because another event occurred, based on how similar they both are. Specifically, if event A is highly similar to event B, then, the probability that event B will occur is higher if event A has occurred. This heuristic can be illustrated the following way: If people are presented with the description of a person and must guess their profession, people will base their decision on the person's characteristics that most resemble the stereotype of that profession. Making decisions based on this heuristic lead people to ignore other influencing factors. The authors enumerate some factors that are typically ignored under the representativeness heuristic (e.g., prior probability, chance, and validity).

The second heuristic described in their article is *availability*. Under this way of reasoning, people judge the frequency or probability of an event based on how easily it is to think about that event. For example, one may judge the frequency of car accidents based on how many occurrences they can think of. The problem with this heuristic is that it is influenced by many biases. There are biases involving the retrievability of occurrences, which is affected by familiarity of the event, by how salient this event is and by whether occurrences are recent or not.

The last heuristic is *adjustment and anchoring*. In some situations, people judge a frequency or probability based on an initial value that is adjusted to make decisions. This initial value is either given to people, or it is the result of an incomplete computation. In their article, the authors gave an example: If people need to estimate the answer to a sequence of multiplications where the numbers are in ascending order and have a limited amount of time to do so, people usually underestimate the answer. The initial value in this case is the first few numbers of the sequence which are smaller than the last few numbers. The problem is that different initial values generate different answers that are all biased toward that initial value. This type of reasoning leads to errors and biases. For example, the anchoring can be based on an incorrect assessment of subjective probability distributions.

There are also other types of heuristics and biases. For example, there's the *gambler's fallacy*, the *base-rate fallacy* and the *conjunction fallacy* (Garfield & Ben-Zvi, 2008). Also, researchers have observed many misconceptions toward specific statistical concepts. For example, people have a difficulty to understand the law of large numbers, trial independence, randomness, correlation/causation and others. A recent study reviewed more exhaustively these heuristics and biases (Husereau et al., 2024).

How Students use Heuristics

Some studies suggest that students do not hold one heuristic in particular but many at different times and even at the same time (e.g., Konold, 1995; Konold et al., 1993). In fact, people can hold two contradicting heuristics for the same problem (Konold, 1995). As part of a study, undergraduates were presented with a two-part coin problem. In the first part, they were asked which head-tail sequence was the most likely. The correct answer was that all the sequences were equally likely. The second part asked participants which sequence was least likely, with the correct answer being that they were all equally unlikely. Konold (1995) reported that 70% of undergraduates answered correctly to the first part of the problem. Out of that 70%, a little over half did not answer correctly to the second part.

The reason of this inconsistency is that students apply different ways of reasoning to the two parts of the problem (Konold, 1995). In the first part, participants who answered correctly did not do so based on the reasoning that all sequences have the same probability of occurrence. Instead, they were unable to rule out any of the response choices. Participants who answered the second part incorrectly did so by choosing the sequence that has too many heads or tails because it was not what they typically see in the population. To put it simply, students went from the outcome approach to the representativeness heuristic. The *outcome approach* is another simplistic way of thinking that Konold (1989) defined as the prediction of the outcome of one single trial depending on the outcome of previous trials. Under this heuristic, trial independence is ignored.

Learning with Heuristics

Heuristics keep students from developing understanding of statistical concepts. Indeed, it has been suggested that most people use heuristics when facing “situations that are less obviously probabilistic or for which the sample space is less tractable” (Konold, 1989, p. 88). However, when situations are obviously probabilistic, people can use formal statistical knowledge. As such, it does not seem to be an inability to use formal statistical reasoning. It is possible that people do not have enough cognitive resources when facing a complex problem. In other words, maybe people use heuristics as a way to make a quick decision without putting too much effort or work into it (Kool et al., 2010). In a statistics course, those effortless decisions are not enough to obtain a good grade considering that heuristics lead to many errors. However, note that the explanation offered here is only conjecture and has not been tested.

Teaching Students with Heuristics

The literature in that area is contradictory. Some researchers argued that statistical reasoning is teachable (Nisbett et al., 1987; Zieffler et al., 2008). Others said that formal statistical reasoning is not teachable because heuristics are too persistent, even after formal training (Kahneman & Tversky, 1972). Some authors are somewhat in between: They reported moderate to small improvement of judgements following training (Adams & Adams, 1958; Lichtenstein et al., 1982).

Sedlmeier and Gigerenzer (1997) looked at this inconsistency in the consideration of sample size. The authors proposed that the law of large numbers – a statistical principle related to sample size – is applied only to certain problems or events. This suggests that people are capable of applying formal statistical knowledge or concepts only to some problems and not others. If the problems are too uncertain with too many factors to consider, people resort to quick simple ways of thinking. This means that educators need to find a way to teach statistics using problems that are obvious and more concrete and gradually increase the abstract nature of problems. In fact, many researchers argued that one of the best ways to get learners to use formal statistical reasoning is to teach with concrete real-world problems (e.g., Bargagliotti et al., 2020; Cousineau, 2020; Garfield & Ben-Zvi, 2008; Harth, 2018; Konold, 1989). Rolling dice and flipping coins are also useful and even though most learners have done those activities often, variations exist (Dunn, 2005).

A recent study proposes a few ways to counter those heuristics (Martín & Valiña, 2023). The authors mentioned a recent study which indicates that training people on cognitive biases, and on the formal reasonings they should use instead depending on the situation (Belton & Dhami, 2021). Doing this, people may be able to intentionally use formal reasonings instead of their heuristics. However, to ensure that this training is effective, people need to be aware of their own heuristics as they often think that they are less affected by those biases (Martín & Valiña, 2023). The authors continued presenting recent studies indicating that training people on certain problems' biases and showing the correct reasonings to use, these people are latter able to correctly solve similar problems (e.g., Boissin et al., 2022a; Boissin et al., 2022b). This research indicates that it may be necessary to teach students about their heuristics and how to override them when teaching statistics.

Another possible way to teach students is by using visualisations (e.g., Caro-Barrera et al., 2023; Forbes, 2012). The role of visuospatial processing in learning is especially true in mathematics, where studies indicate that better spatial processing skills predict better math understanding and performance (Ashkenazi & Velner, 2023; Delage et al., 2022; Ferguson et al., 2015; Sokolowski et al., 2019). This may be the case for statistics as well. Studies A and B of the present thesis explore this hypothesis.

The Role of Spatial Processing Skills

The idea of this thesis comes from a line of research where people's numerical and mathematical skills relate to their skills in mentally manipulating or transforming elements (de Hevia et al., 2008; Dehaene et al., 1993; Delage et al., 2022; Gevers et al., 2006; Gunderson et al., 2012; Lyons et al., 2018). This hypothesis from the field of *Numerical Cognition* stipulates that when a person's spatial skills are strong, the person's numerical and mathematical skills tend to be strong as well. Mathematics anxiety and spatial anxiety are indirect measures of these skills, with lower skill visible in higher anxiety and vice-versa. (e.g., Cheng & Mix, 2014; Cui & Guo, 2022).

The spatial contribution onto mathematics may come from humans' ability to represent numerical magnitudes using a spatial format in nature (Toomarian & Hubbard, 2018). This spatial-numerical associations are manifested through the *Spatial Numerical Association of Response Codes* (SNARC) effect. Even in people who are not mathematically-oriented (e.g., students in non-STEM programs), the spatial representation of number magnitude can be observed (Dehaene et al., 1993). In a task where the researchers asked participants to categorize numbers, participants were faster when the number was larger and the key they had to press was on the right. The opposite results were also observed; participants were faster when the number was smaller and the key they had to press was on the left. This supports the argument that numerical magnitudes are mentally represented in a line from right to left with numbers in ascending order. This is also called the mental number line.

As suggested by the studies above, mathematics seems rooted in spatial skills. It is generally argued that statistics is related to (but different from) mathematics (Baloğlu, 2004; Cruise et al., 1985; Zeidner, 1991). As such, could statistics also be rooted in spatial skills? In other words, could there be a *spatial* line of research, e.g., *Statistical Cognition*? (e.g., Beyth-Marom et al., 2008; Ouyang et al., 2025).

Goals of the Present Thesis

The goal of this thesis is to contribute to the improvement of statistics education by engaging in a new line of research. As said above, the idea of this contribution comes from

Numerical Cognition, and because statistics is related to mathematics, there could be a *Statistical Cognition* that would influence students' understanding of statistics. This line of research would be dedicated to the role of spatial skills in the understanding of and skills or achievement in statistics. The support for this *Statistical Cognition* line of research is beyond the scope of this thesis, though, it is our goal to show that the bridge between these two research fields can be built through the following three studies.

The first study of the present thesis examines the relations between spatial anxiety, mathematics anxiety and statistics anxiety, with gender, social anxiety and trait anxiety as covariates.. The second study goes a step farther by looking at both visuospatial working memory span and spatial problem-solving skills as an indicator of spatial processing skills. Herein, verbal working memory and verbal ability are examined, as well as two anxiety types (spatial, and statistical).

In order to examine the possible existence of statistical cognition (understood as Numerical Cognition), one needs an instrument measuring conceptual understanding of statistics. In order to develop such a test, one needs to know which statistical concepts it will evaluate. The last study of the present thesis examines this prerequisite. Four existing tests of statistics understanding are put together, and statistics instructors are asked to identify the concepts each item is measuring.

Chapter 2: Study A

The Correlates of Statistics Anxiety: Relationships With Spatial Anxiety, Mathematics Anxiety and Gender

Reference

Gibeau, R.-M., Maloney, E. A., Béland, S., Lalande, D., Cantinotti, M., Williot, A., Chanquoy, L., Simon, J., Boislard-Pépin, M.-A., & Cousineau, D. (2023). The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender. *Journal of Numerical Cognition, 9*(1), 16–43. https://doi.org/10.5964/jnc.8199

Contribution

Study conceptualisation, coordination of ethical requests, data collection, data analysis and redaction/revision of the manuscript.

Highlights

- A large sample of 778 university students from 13 different universities and from 2 continents (North America and Europe).
- For better ecological validity, participants are recruited during a statistics class by their statistics instructor.
- Validation of the French version of two scales (*Spatial Anxiety Questionnaire* [SAQ]; *Single-Item Math Anxiety Scale* [SIMA]).
- All three anxiety types (spatial, mathematics, statistics) are related.
- Spatial anxiety directly and indirectly, through mathematics anxiety, predicts statistics anxiety.

Disclaimer

Minor changes were made for the present thesis that are not in the published article to improve the text based on the comments of the thesis committee members.

Abstract

When facing statistical or mathematical operations, people may imagine or visualize the task operations they must do to obtain the result. This study investigates the correlates of statistics anxiety. Considering that statistics anxiety and spatial anxiety have been separately correlated with related constructs (e.g., mathematics anxiety, academic performance, etc.), the possibility that spatial anxiety plays a role in statistics anxiety is explored. To examine this question, 778 students in a social or health sciences program, enrolled in a – often mandatory – statistics course from Canadian, French, and Belgian universities completed an online survey. The results show moderate to strong positive correlations between all three types of anxiety (spatial, mathematics, and statistics). In addition, a mediation analysis reveals the intermediate role played by mathematics anxiety in the relationship between spatial and statistics anxieties. Nonetheless, the direct link from spatial anxiety to statistics anxiety is non-negligible in the model. Finally, the results also indicate that women report higher levels of statistics anxiety, which may be partly explained by their higher level of spatial anxiety. This study has implications for statistics education, where the anxiety toward statistics could be decreased by treating spatial anxiety.

Keywords: statistics anxiety, spatial anxiety, statistics education, mathematics anxiety

Knowing how to appraise data, interpret them, and stay critical when faced with quantitative information are constantly solicited abilities not only at the professional level but also in daily life. For example, every day many people encounter quantitative information and statistical concepts such as probabilities (e.g., meteorology) and frequencies (e.g., epidemiology). Shaughnessy (2010) claimed that the most important cognitive ability is to be able to read and interpret statistical and probabilistic information, and to be able to make inferences and decisions based on that information. Yet, many studies have revealed a widespread discomfort regarding quantitative information (Paechter et al., 2017) occurring "as a result of encountering statistics in any form and at any level" (Onwuegbuzie et al., 1997, p. 28), a state called statistics anxiety. The present study aims to explore a possible influencing factor of *statistics anxiety*, namely, *spatial anxiety*. This last phenomenon is the anxiety towards activities requiring spatial skills (e.g., driving in an unfamiliar city).

Statistics Anxiety

While quantitative data are presented to people in different contexts, many have difficulties understanding their meaning and implications (Konold, 1995). Kahneman (2011) argued that human beings are not particularly good *homo statisticus*, which refers to humans being statistically competent (Béland et al., 2016). Consequently, at least in humanities and social sciences university programs – in which statistics courses are often mandatory –, many students feel uncomfortable with that subject (Cousineau & Harding, 2017). The presence of statistics anxiety has been amply investigated (e.g., Benson, 1989; Birenbaum & Eylath, 1994; Cruise et al., 1985; Cui et al., 2019; DeVaney, 2010; Onwuegbuzie et al., 1997; Zanakis & Valenzi, 1997), and statistics have been found to generate high levels of anxiety in university students (Baloğlu, 2003). Baloğlu (2003) looked at statistics anxiety mean levels separated by age and gender. The results indicate that men younger than 21 years-old reported a mean level of 132.41/255, women of the same age reported a mean level of 131.56/255; men between 22-26 years-old reported a mean level of 123.73/255, women of the same age reported a mean level of 125.22/255; men older than 27 years-old reported a mean level of anxiety of 114.15/255, women of the same age reported a mean level of 133.33/255.

While there is consensus in the literature that statistics anxiety is a multidimensional construct (e.g., Cruise et al., 1985), there is not yet agreement on its exact nature and components. Cruise et al. (1985) described six features: the worth of statistics, test and class anxiety, interpretation anxiety, computational self-concept, anxiety of asking for help and anxiety of statistics teachers. Zeidner (1991), on the other hand, reported only two facets: test anxiety and content anxiety. Onwuegbuzie et al. (1997) suggested four facets: instrumental anxiety, content anxiety, interpersonal anxiety, and failure anxiety. Finally, Vigil-Colet et al. (2008) proposed three factors: test anxiety, anxiety of asking for help and interpretation anxiety. This last operationalization was used herein as the number of items is more consistent through the subscales and because the scale is shorter, compared to Cruise et al.'s instrument. Chew et al., (2018) also indicate that there are two parts to the STARS; one that measures statistics anxiety and a second that measures attitudes towards statistics. Considering that the present study examine only statistics anxiety, Vigil-Colet et al.'s (2008) framework was used instead. In other words, the *Statistical Anxiety Scale* (SAS; Vigil-Colet et al., 2008) was used herein instead of the *Statistical Anxiety Rating Sacle* (STARS; Cruise et al., 1985).

To overcome this lack of consensus, some investigators tried to understand what statistics anxiety is by examining how it relates to and differs from other constructs, and how people's level of statistics anxiety varies. One factor commonly examined to explain individual differences in multiple domains is gender (e.g., Baer & Kaufman, 2008). However, the investigation of gender does not provide consistent effects across studies. Some researchers observed a difference between men and women, where women report higher statistics anxiety (e.g., Lalande et al., 2019; McIntee et al., 2022), and others concluded that there is no such difference (e.g., Baloğlu et al., 2011; Bui & Alfaro, 2011; DeMaria Mitton, 1987; Hsiao & Chiang, 2011; Mandap, 2016). The effect of statistics anxiety on performance may especially be detrimental to women as they often (but not always) report higher levels of a variety of anxieties (Casey et al., 1995, 1997; Eduljee & LeBourdais, 2015; Lawton, 1994). Though, this gender effect may be partly explained by environmental and social factors, considering that women are taught to talk about their emotions (Brebner, 2003; Frenzel et al., 2007). Other research suggested that increased statistics anxiety is associated with decreased interest towards the subject, lower mathematics self-concept, lower mathematical ability and higher trait anxiety (Bourne, 2018; Macher et al., 2012). Statistics anxiety is also positively related to procrastination in statistics courses (Lalande et al., 2019; Onwuegbuzie, 2004; Walsh & Ugumba-Agwunobi, 2002) and to mathematics anxiety and mathematics achievement (Barroso et al., 2021; Zeidner, 1991), mathematics anxiety referring to fear, tension, and apprehension when faced with mathematics (Ashcraft, 2002). In addition, it has been proposed that statistics anxiety and mathematics anxiety affect academic performance because of its disruptive impact on student learning (Onwuegbuzie & Wilson, 2003; Primi & Chiesi, 2018). Indeed, statistics anxiety is an important predictor of performance in statistics courses (Cantinotti et al., 2017; Fullerton & Umphrey, 2016). A recent study indicate an opposite result, where performance in a statistics course predicts statistics anxiety (Mackinnon et al., 2025).

Against this background, a question remains unanswered; why do statistics induce anxiety in university students? In other words, what are the factors that cause students to experience anxiety when facing statistics? The answer to this question might come from an unexpected direction.

Spatial Anxiety

Much research suggested that spatial skills are associated with mathematics abilities (e.g., (Cheng & Mix, 2014; de Hevia et al., 2008; Martinez, 1987; Mix et al., 2016; Mix & Cheng, 2012; Tosto et al., 2014). This association is observed in children as young as 3- and 4-year-old (Verdine et al., 2014). It is generally found that poor spatial skills are associated with weaker mathematical performance and achievements (Rotzer et al., 2009; Verdine et al., 2017). A longitudinal study looked at the relation between spatial skills and mathematics skills. The authors reported an association between spatial skills and mathematics performance at their first time point (children 7 to 9 years old), and reported that mathematics performance at their second time point (the next school year) predicts spatial skills at their last follow-up (year 3; Geer et al., 2019). These findings suggest a loop of influence where early spatial skills predict mathematics skills, which predict later spatial skills.

Further evidence of the influence of spatial skills can be found in the Science, Technology, Engineering, and Mathematic fields (STEM; Atit & Rocha, 2020; Lyons et al., 2018). Indeed, Eme and Marquer (1999), and others, proposed that individual differences in STEM fields share a common root: spatial skills. For example, Casey et al. (1997) found that mental rotation, a spatial skill, is a predictor of performance on the mathematics scale of the Scholastic Aptitude Test (SAT-M). Furthermore, pre-teens' and teens' spatial skills predict whether or not they choose STEM programs at the post-secondary level and to work in STEM professions (Atit & Rocha, 2020). It is argued that spatial skills act as a gateway towards STEM fields, where learners' spatial skills help them achieve expertise (Hambrick & Meinz, 2011; Stieff, 2004; Uttal & Cohen, 2012). There is very little agreement on the specific types of spatial skills (Atit et al., 2022; Atit & Rocha, 2020). However, many researchers agreed that typical tasks in STEM are based on skills such as finding a new perspective of an object by mentally "moving" it (Atit et al., 2020, 2022).

As weaker spatial skills could induce anxiety, the relatively new construct of spatial anxiety has been proposed. Indeed, a recent study found that spatial anxiety (described below) moderate the relation between spatial skills and numerical skills (Ouyang et al., 2022). Relatedly, Geer et al. (2024) explained the relation between spatial anxiety and spatial skills. The authors argued that there are two possible mechanisms at play. In the first one, experience avoidance plays a role, where people avoid situations that make them feel negative sensations like anxiety (Berman et al., 2010; Geer et al., 2024). As such, people who avoid tasks or situations that require spatial skills because of their anxiety do not develop those skills (Geer et al., 2024). In turn, their underdeveloped skills lead to domain-specific anxiety, like spatial anxiety. This also suggests a loop of influence. In the second possible mechanism, the domain-specific anxiety requires too much cognitive resources impacting the effectiveness of working memory while doing a task.

First introduced by Lawton (1994), spatial anxiety is characterized by a feeling of nervousness or stress when facing spatial operations or activities. Spatial operations include tasks that require mentally visualizing, rotating or transforming spatial and visual information (Gardner, 1993). Lawton (1994) developed a tool to measure spatial anxiety. However, this measure focused solely on navigation anxiety. Malanchini et al. (2017) also developed a scale to measure spatial anxiety, but the dimension of imagery does not seem to be included. Finally, Uttal et al.'s (2013) proposed a 2 × 2 framework of spatial skills based on the work of Newcombe and Shipley (2015). Lyons et al. (2018) tested this framework and identified three distinct kinds of spatial skills: mental manipulation, navigation, and imagery. They also developed a scale to measure anxiety related to these spatial domains. Mental manipulation refers to the ability to manipulate information or objects in one's mind; navigation is the ability to orient oneself in the environment; imagery is the ability to visualize elements in detail.

There is evidence linking navigation abilities to spatial anxiety, with increased anxiety being associated with decreased performance on a navigation task (i.e., the sense of direction; Kremmyda et al., 2016). It is known that spatial skills are extremely solicited in everyday life (Vasilyeva & Lourenco, 2010), whether it is for navigating an area, visualizing the interior and exterior design of a house, or even deciding whether a car fits into a parking spot.

Spatial anxiety can be seen in children as young as 6-8 years of age (Lauer et al., 2018; Ramirez et al., 2012). This type of anxiety has long-lasting consequences. For example, when elementary-school teachers are anxious about spatial tasks, their students develop their spatial proficiency less over the school-year relative to their peers with teachers who are low in spatial anxiety (Gunderson et al., 2013). In turn, less developed spatial skills limit mathematics skills and decrease entry into STEM programs, as mentioned above.

Gender differences have also been examined both in spatial anxiety and spatial skills. Some studies report higher levels of spatial anxiety in women and girls (Alvarez-Vargas et al., 2020; Lauer et al., 2018, while others do not observe this difference, Lemieux et al., 2019) Additionally, some studies indicated a gender difference in specific spatial skills (e.g., mental rotation and navigation; Jordan et al., 2002; Ratliff et al., 2009; Schmitz, 1999; Wong, 2017). A recent study suggested that spatial anxiety mediates the relation between gender and mental rotation performance (Alvarez-Vargas et al., 2020).

Spatial anxiety is related to mathematics anxiety, and those who have lower spatial skills are higher in mathematics anxiety (Douglas & LeFevre, 2018; Ferguson et al., 2015; Maloney et al., 2012). Sokolowski et al. (2019) reported a similar finding where spatial anxiety mediates the relation between gender and mathematics anxiety. While it may seem odd that mathematics anxiety would be linked to the domain of spatial processing, there is an extensive literature linking mathematics and space. Two separate series of studies have examined either statistics anxiety or spatial anxiety using similar correlates. It is known that both constructs are negatively related to mathematics aptitudes and positively related to mathematics anxiety (Ferguson et al., 2015; Fullerton & Umphrey, 2016; Malanchini et al., 2017; Zeidner, 1991). Statistics and spatial skills are also both positively associated with mathematics abilities (Casey et al., 1997; Douglas & LeFevre, 2018; Ferguson et al., 2015). Moreover, both are correlated with performance in several activities or evaluations, like statistics exams or performance in a navigation task (e.g., Hund & Minarik, 2006; Macher et al., 2012; Onwuegbuzie & Wilson, 2003). Considering those similar correlations, there may exist a relation between statistics anxiety and spatial anxiety.

Other research has suggested that if spatial skills are related to mathematics anxiety, then they could also be related to statistics anxiety (Ferguson et al., 2015). Testing this hypothesis is the main purpose of the present study. Two related measures were included in the study as covariates. First, it has been established that trait anxiety is moderately and positively correlated with mathematics anxiety and statistics anxiety (Malanchini et al., 2017). Therefore, trait anxiety was measured. Second, the anxiety of asking for help in statistics represents an anxiety that may be social in nature. As such, a measure of social anxiety was also included.

The first goal of the present study is to determine the strength of the relation, if one exists, between statistics anxiety and spatial anxiety while taking into consideration other types of anxiety that may have an influence (i.e., social anxiety and trait anxiety). It is hypothesized that this relationship is partially mediated by mathematics anxiety. The reason for this is because it is observed early in people's schooling (sometimes as early as 8-9 years old; Sorvo et al., 2019) and because STEM (which specifically includes mathematics) is highly valued in primary and secondary school education, contrary to statistics (Banerjee, 2016). A subsidiary goal of this work is to validate the French versions of three scales that were translated for the present study.

Method

Participants

The participants were French-speaking students recruited from Universities in Canada, France, and Belgium. They were undergraduates in social, human or health sciences programs, and they were enrolled in a statistics course at the time of the study. A total of 16 universities were solicited to reach at least 500 participants. This number was chosen a priori (gauging the effect sizes we previously obtained in related studies, e.g., Cantinotti et al., 2017; Lalande et al., 2019, and McIntee et al., 2022, and informally assuming that the added measures would not be as strong). After the experiment was completed, we ran a power analysis. It suggests that for comparing standardized regression coefficients with differences of about 0.1, the targeted sample size of 500 would yield statistical power of 85% (with the sample size that was obtained, this figure rises to 95%). Further, according to Comfrey and Lee (1992), in order to do a path analysis, a sample of 500 participants is sufficient. No exclusion criteria were applied in this study to maximize the validity, or representativeness, of the sample. All participants were recruited in the statistics class using a static one-page slide advertisement which may have been accessible in their course material repository as well.

Materials

The questionnaire was composed of six different scales. These scales were (1) a French translation of the Spatial Anxiety Questionnaire (SAQ; Lyons et al., 2018), (2) a French translation of the Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003), (3) the French version of the Statistical Anxiety Scale (SAS-F; Cantinotti et al., 2017), (4) the validated French version of the Liebowitz's Social Anxiety Scale (LSAS-F; Yao et al., 1999), (5), the validated French version of the State-Trait Anxiety Inventory by Speilberger (STAI-Y-F; Gauthier & Bouchard, 1993) and lastly, (6) a French translation of the Single-Item Math Anxiety Scale (SIMA; Núñez-Peña et al., 2014). The SAQ, the AMAS and the SIMA were translated in-house for the present study because a French version was not otherwise available. The last author of the present article translated all three scales. Then, three revisions of the translated scales were made by other co-authors. The final translation can be found in Appendix A.

After providing informed consent, participants completed a set of sociodemographic questions. Next, participants completed items from the six scales, divided into 13 blocks. The first four blocks, presented in a randomized order to avoid order effects, were composed of seven items from the SAS, eight from the SAQ and from the LSAS-F (all selected randomly) and the nine items of the AMAS. The blocks five to nine, also in a randomized order, presented another seven items from the SAS, another eight from the SAQ, eight items from the LSAS-F and from the STAI-Y, and the unique SIMA item. Finally, the last four blocks (10 to 13; order randomized) were composed of the remaining seven items of the SAS, the remaining eight of the SAQ and the LSAS-F and the remaining twelve items of the STAI-Y.

Questionnaires

Spatial Anxiety

To measure this construct, a French translation of the Spatial Anxiety Questionnaire (SAQ) was administered (Lyons et al., 2018). Based on 24 items, the SAQ has three factors: Mental Manipulation (e.g., "asked to imagine the 3-dimensional structure of a complex molecule using only a 2-dimensional picture for reference"; M), Navigation (e.g., "asked to find your way back to your hotel after becoming lost in a new city"; N) and Imagery (e.g., "asked to recall the shade and pattern of a person's tie you met for the first time the previous evening"; I). Participants answer on a five-point scale ("not at all" to "a lot") with scores of 0 to 4 respectively. The range of possible scores is 0-96, with higher scores indicating higher anxiety. Cronbach's alpha suggested a good to excellent reliability for all subscales (M: .92, N: .91 and I: .86) in the original study. With the present sample, the Cronbach's alpha and McDonald's omega suggest an excellent reliability for the total scale (both .93). The authors examined the measure's convergent validity using correlations between each subscale and its corresponding skill. They reported moderate-to-strong correlations, indicating good convergent validity.

Mathematics Anxiety

Mathematics anxiety was measured using the Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003). This measure was translated into French for the present study. The AMAS is a 9-item questionnaire which asks participants to rate how anxious they would feel in a variety of situations (e.g., "having to use the tables in the back of a math book" and "listening to a lecture in math class") on a five-point scale ranging from "low anxiety" (1) to "high anxiety" (5). The range of possible scores is 9-45, with higher scores indicating higher anxiety. Two components have been identified for this scale, learning mathematics anxiety (LMA; e.g., "starting a new chapter in a math book") and mathematics evaluation anxiety (MEA; e.g., "thinking about an upcoming math test one day before"). Good internal consistency was observed in the original study with an alpha of .90 for the total score and alphas of .85 and .88 for the learning and evaluation subscales respectively. With the present sample, the Cronbach's alpha and McDonald's omega suggest an excellent reliability of the total scale (both .90). The validity of this measure was not tested. However, it was observed in a previous study that the AMAS is positively correlated with the *Test Anxiety Inventory*. The correlation is .54 for the whole scale, .34 for the LMA, and .58 for the MEA (Primi et al., 2014). These results suggest good convergent validity.

Statistics Anxiety

This variable was measured using the Statistics Anxiety Scale (Vigil-Colet et al., 2008). The French version was validated by Cantinotti et al. (SAS-F; 2017). This questionnaire is composed of three subscales: evaluation anxiety (e.g., "doing the final exam in a statistics course"), anxiety of asking for help (e.g., "asking a question to the professor concerning my difficulty to understand the material"), and interpretation anxiety (e.g., "trying to understand the lottery's probabilities"). Each subscale is measured by eight items. Participants answered on a five-point scale which ranged from "no anxiety" (1) to "a lot of anxiety" (5). The range of possible scores is 21-105, with higher scores indicating higher anxiety. Note that, in the French version, three items from the original version did not contribute enough to the confirmatory factor structure (i.e., "copying a mathematics example while the professor explains it", "seeing a classmate studying a results' table from a problem they worked on" and "going to a statistics exam for which I didn't have time to study"). These items were eliminated from the questionnaire, so the French version had 21 items, showing excellent internal consistency with a

McDonald's omega of .95 (.85, .98 and .83 for the three subscales). With the present sample, the Cronbach's alpha and McDonald's omega suggest an excellent reliability of the total scale (.95 and .94, respectively). The authors also looked at the measure's divergent validity. The SAS has a negative correlation with a self confidence instrument. ($r = -.42$). This indicates that the SAS has good discriminant validity. Convergent validity was examined by a correlation between the scale and participants' final grade in their statistics course, and by a factor analysis revealing a 3-factor model.

Social Anxiety

The French version of the Social Anxiety Scale by Liebowitz (LSAS-F) was administered (Heeren et al., 2012). Participants answered on a four-point scale ranging from "none" (0) to "severe" (3). The range of possible scores is between 0 and 72, with higher scores indicating higher anxiety It is a 24-item scale composed of two subscales, one assessing anxiety during social performances (e.g., "eating in public") and another assessing anxiety during social interactions (e.g., "meeting new people"). In completing this scale, participants rated how much anxiety they would experience in various situations, e.g., "being the center of attention". Whereas reliability was not quantified in these original articles, Beard et al. (2011) reported that this scale had excellent internal consistency with a Cronbach's alpha of .95 for both subscales. With the present sample, the Cronbach's alpha and McDonald's omega suggest an excellent reliability of the total scale (both .93). Convergent and discriminant validity was tested in Heimberg et al (1999). Convergent validity was tested using correlations with other clinical instruments for social anxiety and avoidance (with the *Social Interaction Anxiety Scale* $r = .73$; with the *Social Phobia Scale* $r = .61$; with the *Fear of Negative Evaluation Scale* $r = .49$; with the *Social Avoidance and Distress Scale* $r = .63$; with the social phobia subscale of the *Fear*

Questionnaire $r = .63$; and with the *Clinician's Severity Rating* of the ADIS-R or ADIS-IV-L interview $r = .52$). The second type of validity was tested using correlations with a general anxiety measure and a depression measure (with the *Hamilton Anxiety Scale* $r = .48$; with the *Beck Depression Inventory* $r = .39$; with the *Hamilton Rating Scale for Depression* $r = .52$). The authors conclude that the SAS has good convergent and discriminant validity.

Trait Anxiety

This variable was measured using the trait anxiety subscale from the State-Trait Anxiety Inventory (STAI-Y-F) created by Spielberger et al. (1971) and translated into French and validated by Gauthier and Bouchard (1993). Cronbach's alpha was .90 in that validation study. In the STAI-Y-F, participants answered 20 items assessing their trait anxiety on a four-point scale ranging from "almost never" (1) to "almost always" (4). The range of possible scores is 20-80, with higher scores indicating higher anxiety. Examples of items include: "I feel safe and without worries" and "I easily make decisions". With the present sample, the Cronbach's alpha and McDonald's omega suggest an excellent reliability of the total scale (both .92). Note that some items are reversed coded, for example "I feel like difficulties are accumulating – I can't overcome them". The scale's validity was tested in Knowles and Oulatunji (2020). Convergent validity was tested using correlations with other anxiety measures (the *Penn State Worry Questionnaire*; the *Obsessive Compulsive Inventory – Revised*; the *Yale- Brown Obsessive-Compulsive Scale*; the *Davidson Trauma Scale*; the *Acute Stress Disorder Interview*; the *Liebowitz Social Anxiety Scale*; the *Social Interaction Anxiety Scale*; the *Social Phobia Inventory*; and the anxiety subscale of the *Symptoms Rating Scale for Depression and Anxiety*). Discriminant validity was tested using correlations with measures of depression (the *Beck Depression Inventory*; and the *Beck Depression Inventory-II*). While individual correlations are not presented because this study is a meta-analysis, the mean correlation with the anxiety measures (r = .59, 95%CI = [.49, .68]) and with the depression measures (r = .61, 95%CI = [.53, .68]) are given. These results indicate that there might be a discriminant validity issue with depression.

The Single-Item Math Anxiety Scale (SIMA)

The last scale was administered as a second measure of mathematics anxiety. It was included in the present study only as an opportunity to translate it in French and validate that version. The scale is composed of only one item – "On a scale from 1 to 10, how math anxious are you?", with "10" being the highest level of anxiety.

Procedure

Ethics approval from all the collaborative universities were obtained prior to data collection. The first page of the questionnaire was the consent form. Participants gave informed consent when answering the question: "Do you consent to participate in this study?" If they chose not to consent, they were redirected to the end of the questionnaire. All the data was obtained in class prior to the first COVID-19 confinement. Participation in this study required answering questions presented in an online questionnaire which took approximately 20 minutes to complete. They could complete the questionnaire on their own time. No compensation was given to the participants, although they had the option of entering a draw to win a CAN $50 Amazon gift card. One gift card for every 50 participants was drawn in January and April 2020.

For some universities the course schedule was publicly accessible. As such, each statistics professor was personally contacted via email to ask if they could share our advertisement materials for the study. Other universities do not have their course schedule publicly available. As such, the psychology department of those universities were contacted via email. Once the department shared their statistics professors' information, we personally contacted them via email. The students were told that we needed their help to understand and decrease the impact of statistics anxiety.

Planned Analyses

After screening the data set for outliers and missing data, the normality of distributions and descriptive statistics were computed to describe the sample. To ease comparisons, all the scale and sub-scale totals were transformed into a score ranging from 0 to 100 by subtracting the minimum possible score for that scale or subscale from the observed score, dividing by the possible range of scores and multiplying by 100. In other words, percentage scores were computed. Next, to determine if the responses from the two continents (North America and Europe) were equivalent, a multiple-group invariance analysis was produced (Meredith, 1993). To do so, the Lavaan R package was used (Rosseel, 2012). Based on a Confirmatory Factor Analysis (CFA), this strategy investigates how the total scores of all the scales are related to a single broad construct, here anxiety in quantitative methods. More critically, it assesses whether the factor structure is equivalent across continents. Invariance between groups will allow the analysis of the continents together instead of separately.

Also, because three of the scales were translated into French for the present study and have not been validated in that language in previous studies, CFAs were carried out to examine if the scales have the same factor structure in French as they do in their original language. CFAs were used for scale validation (Cappelleri et al., 2014; Levine, 2005). Cronbach's alpha and McDonald's omega are also reported.

Lastly, and more central to our investigation, Pearson's correlations were calculated in SPSS between all the scales' scores to examine if, and to what extent, they are associated with one another. Mediation models and multiple regressions, computed with the Lavaan R package, follow to investigate more finely the role of spatial anxiety and mathematics anxiety onto statistics anxiety (MacKinnon et al., 2007). The goals are to understand how spatial anxiety and statistics anxiety are associated, how spatial anxiety influences the levels of statistics anxiety through mathematics anxiety, to test whether this association is still present when controlling for other variables and to understand how other variables influence the association. Following Cousineau's (2020) recommendation, all the raw descriptive statistics are reported with one decimal and all the standardized ones with two decimals. All the statistics are reported with three significant digits. This means one decimal for the descriptive statistics and two decimals for the non-standardized regression coefficients and the standardized effect sizes.

Results

Screening of the Data

In total, 950 participants went to the survey website. Of these, 780 participants went through the study up to the end, but two returned only blank responses and were removed. Of the remaining 778 participants, 377 were enrolled in Europe and 401 in North America, a fairly equal number of participants from both continents. Regarding age, North American students tend to be older (mean age of 22.8 years) than European students (mean age of 21.6 years; both with a standard deviation of 5.6).

Twenty-six items had no missing response, 36 items had one missing response, 26 had two missing responses, five had three missing responses and 6 had four missing responses, for a total of 129 missing responses over 77,022 items (99 items per participant times 778 participants) or less than two in a thousand missing responses. The distribution of the missing responses was random. The Little test was used to examine patterns in the missing responses.

The missing responses were not included when computing the scales' means. No scales or subscales had more than 3 responses missing per participants. Regarding the SIMA, two missing responses imply that this score is not available for two participants. Finally, no multivariate outlier was detected.

Establishing Equivalence Between Continents

Before going any further, it is necessary to determine whether the two continents had similar patterns of responses. To that end, a multiple-group invariance analysis was run in two steps. First, the way the total scores of all six scales were related to a single construct was estimated. The CFA shows standardized factor loadings of .93 for SAS-F, .86 for AMAS, .61 for SAQ, .80 for SIMA, .53 for LSAS and .53 for STAI-Y. The fit indices are poor (CFI = .82, TLI = .70, RMSEA = .26 and SRMR = .11) because it is not suggested that a single factor underlies all six scales. This CFA serves as a baseline to evaluate group invariance.

Second, four CFA models were used to measure group invariance between North Americans and Europeans (Xu, 2012). In Model 1, there is no restriction in the model parameters. This model, the least parsimonious, has 36 free parameters. In Model 2, to assess metric invariance, the factor loadings are constraint to be equal across continents, resulting in 31 free parameters. In Model 3, testing for scalar invariance, intercepts and factor loadings are constraint to be equal across continents, leading to 26 free parameters. In Model 4, testing strict or full invariance, factor loadings, intercepts and residual variances are set to be equal across continents, requiring 20 free parameters. The fit of these models are shown in Table 1.

If there are some small differences, there is overall not a huge decrement in fit as constraints are added. Both AIC and BIC favored model 2, suggesting that the factor loadings do not differ between continents. However, model 3 imposing equal intercepts was poorly supported, which suggests that the continents differ on the mean scores.

Table 1

Multiple Group Invariance Analysis: Model Fit Measures

Models	Number of free parameters	Log-likelihood	CFI	TLI	RMSEA	SRMR	AIC	BIC
Model 1	36	−19,687.2	.824	.707	.262	.093	39,446.4	39,613.9
Model 2	31	−19,689.1	.825	.771	.231	.096	39,440.3	39,584.5
Model 3	26	−19,715.2	.807	.794	.220	.102	39,482.4	39,603.4
Model 4	20	−19,727.6	.800	.824	.203	.105	39,495.3	39,588.3

Note. Model 1 is the baseline; Model 2 is the metric invariance model; Model 3 is the scalar invariance model and Model 4 is the strict or full-invariance model.

Table 2 indicates the mean scores broken down by continents. The most different scores are social anxiety (LSAS) and trait anxiety (STAI-Y) which are covariates in subsequent models. Thus, even if we do not have full invariance, we have metric invariance, i.e., invariance at the level of the regression coefficients which are utilized in the subsequent pages. For example, the factor loadings in Model 2 are for North America, .92, .84, .63, .78 .52, and .51; compared to .93, .88, .61, .82, .54, and .55 in Europe, which are all within ± .02 of the baseline solution provided in Model 2.

Table 2

Means on Scales' Total Scores Subdivided by Continents

	SAQ	AMAS	SAS-F	LSAS-F	STAI-Y-F	SIMA
North America	40.9	45.9	46.5	44.7	46.9	54.4
	[39.6, 42,2]	[44.3, 47.5]	[45.0, 48.0]	[43.4, 46.0]	[45.5, 48.3]	[52.3, 56.5]
Europe	38.1	44.7	49.5	48.9	52.3	55.2
	[36.9, 39.5]	[43.1, 46.3]	[48.0, 51.0]	[47.6, 50.2]	[50.9, 53.7]	[53.1, 57.3]
Cohen's d_p	0.15	0.05	−0.14	−0.22	−0.28	−0.03
	[0.01, 0.29]	[−0.09, 0.19]	[−0.28, −0.00]	[−0.36, −0.08]	[−0.42, −0.14]	[−0.17, 0.11]

Note. Cohen's d_p and its confidence intervals for between-groups are computed as per Goulet-Pelletier and Cousineau (2018).

Descriptive Statistics and Reliability

Regarding gender, 84.5% (660 participants) were women, 14.5% (113 participants) were men, 0.8% (4 participants) answered “other or prefer not to say” and one did not respond. The gender distribution is nearly identical in both continents (e.g., 13% of men in Europe vs. 16% in North America). The higher proportion of women is to be expected, as women make up the majority of the student population in psychology and other social sciences (Odic & Wojcik, 2020), where most participants were recruited. Age of respondents ranged from 17 to 61 years old (55 years old in North America). The average age was 22.1 years old (SD = 5.6).

For all scales, the higher the score, the more severe the anxiety (the STAI-Y was built so that low scores mean higher anxiety; we reversed this scale using 100 minus the total score on 100). Descriptive statistics for the 6 questionnaires are presented in Table 3. We also provide in Table 3 the 95% confidence intervals where relevant (Harding et al., 2014). As seen from the shape statistics (skewness and kurtosis), the total scores are not normally distributed, but the deviations are not severe. Therefore, the data were not further normalized.

Cronbach's alpha and McDonald's omega were computed and are reported in the last two lines of Table 3. McDonald’s ω (Béland et al., 2017) were estimated using the HA algorithm from Hancock and An (2020) using Hayes and Coutts (2020) implementation. For all scales, the reliability is excellent and matches reliability reported in the original publications, the weakest Cronbach’s alpha being .90 for the AMAS, identical to the value reported in the original study. As a safeguard, we also computed McDonald's omega based on a single dimension, but the reliabilities are identical to two decimals.

There is a strong gender difference regarding statistics anxiety, as women score on average 49.4 out of 100 whereas men score on average 38.3 (standard deviations of 20.3

and 21.3 respectively), a difference of 18.0 with 95% confidence interval of the difference of [−15.2, −7.0]; hereafter, the square brackets will be used to denote 95% confidence intervals of the difference. This difference is strongly significant (b = 2.12, SE = 1.13, $p < .01$). However, subsequent analyses will suggest that this effect may have its origin in spatial anxiety. On that variable, women score 41.0 out of 100 on average, compared to men who score 30.7 on average (standard deviations of 17.8 and 17.5, respectively, again with a strongly significant difference, b = 9.77, SE = 1.77, $p < .01$).

Table 3

Descriptive Statistics for the Six Questionnaires' Total Scores Reported on a Score Ranging from 0 to 100 where Higher Scores Represent Higher Levels of Anxiety.

	SAQ	AMAS	SAS-F	LSAS-F	STAI-Y-F	SIMA
Mean	39.6	45.3	47.9	46.7	49.5	54.8
	[38.3, 40,9]	[43.7, 46.9]	[46.5, 49.4]	[45.4, 48.1]	[48.1, 50.9]	[52.7, 56.9]
Minimum	0	0	1	0	0	0
Maximum	98	100	100	100	95	100
n	778	778	778	778	778	776
SD	18.1	22.4	20.9	19.1	19.4	29.4
	[17.3, 19.1]	[21.4, 23.6]	[19.9, 21.9]	[18.2, 20.1]	[18.4, 20.4]	[28.0, 30.9]
Skewness	0.29	0.14	0.14	0.16	−0.07	−0.27
	[0.12, 0.46]	[0.02, 0.24]	[0.02, 0.25]	[−0.02, 0.33]	[−0.24, 0.10]	[−0.44, −0,10]
Kurtosis	−0.28	−0.71	−0.56	−0.46	−0.65	−1.10
	[−0.60, 0.09]	[−1.03, −0.34]	[−0.88, −0.19]	[−0.78, −0.09]	[−0.97, −0.28]	[−1.41, −0.73]
Cronbach α	.93	.90	.95	.93	.92	NA
	[.93, .94]	[.89, .91]	[.94, .95]	[.92, .94]	[.91, .93]	
McDonald ω	.93	.90	.94	.93	.92	NA
	[.92, .94]	[.88, .90]	[.94, .95]	[.92, .94]	[.91, .93]	

Note. SAQ is the *Spatial Anxiety Questionnaire*, AMAS is the *Abbreviated Math Anxiety Scale*, SAS-F is the French version of the *Statistics Anxiety Scale*, LSAS-F is the French version of the *Liebowitz's Social Anxiety Scale*, STAI-Y-F is the French version of the trait anxiety subscale of the *State-Trait Anxiety Inventory* and SIMA is the *Single-Item Math Anxiety* scale. Numbers between braces are 95% confidence intervals. The whole descriptive statistics broken down by continents are available in the supplementary material at https://osf.io/jpv5n/

Validating French Versions of the Scales' Scores

As mentioned in the Method section, three scales were translated into French for the current study (SAQ, AMAS and SIMA). Confirmatory factor analyses were performed on the SAQ and the AMAS. The same structure described in the original articles was tested to examine if it fits the data of the current sample. Factor loadings are presented in Table 4 and Table 5 for the SAQ and the AMAS respectively. The loadings vary from .54 to .84 in Table 4, and from .69 to .83 in Table 5. Fit indices were excellent for both translated questionnaires (for the SAQ-F: CFI = .96, TLI = .95, RMSEA = .05 and SRMR = .05; for the AMAS-F: CFI = .98, TLI = .98, RMSEA = .06 and SRMR = .03). To validate the French translation of the SIMA, the instrument was correlated with the AMAS. The result (r = .79) indicates excellent content validity of the questionnaire, congruent with what is reported in the original article (r = .77 for the correlation with the sMARS, a questionnaire from which AMAS was adapted). The two CFA models along with the coefficients are provided in the Supplementary Materials.

Table 4

Factor Loadings for the SAQ-F

	Spatial Anxiety Questionnaire		
	Mental Manipulation	Navigation	Imagery
1	.81* [.78, .84]		
2	.78* [.75, .81]		
3	.77* [.74, .81]		
4	.75* [.71, .78]		
5	.78* [.75, .81]		
6	.72* [.68, .75]		
7	.70* [.66, .74]		
8	.76* [.72, .79]		
9		.83* [.81, .86]	
10		.82* [.79, .85]	
11		.82* [.79, .85]	
12		.78* [.75, .81]	
13		.84* [.81, .86]	
14		.83* [.80, .85]	
15		.80* [.77, .83]	
16		.54* [.49, .59]	
17			.72* [.68, .76]
18			.78* [.75, .81]

19	.70* [.66, .75]
20	.65* [.60, .70]
21	.60* [.55, .65]
22	.57* [.52, .62]
23	.64* [.58, .67]
24	.61* [.56, .66]

Note. * $p < .001$. Numbers between brackets are the 95% confidence intervals.

Table 4

Factor Loadings for the AMAS-F

	Abbreviated Math Anxiety Scale	
	Learning	Evaluation
1	.75* [.72, .79]	
2	.74* [.71, .78]	
3	.76* [.73, .80]	
4	.80* [.77, .83]	
5	.79* [.76, .82]	
6		.78* [.74, .81]
7		.83* [.80, .86]
8		.69* [.65, .74]
9		.75* [.70, .78]

Note. * $p < .001$. Numbers between brackets are the 95% confidence intervals.

The excellent fit of both CFAs and the strong correlation between the SIMA and the AMAS indicate that, the translations, hereafter labeled SAQ-F, AMAS-F and SIMA-F, were appropriate measures for spatial anxiety and mathematics anxiety, respectively, in a francophone population.

Associations Between the Variables

Correlations Between Scales

Pearson's correlations were computed between each scale's total score. The correlation matrix is presented in Table 6. As anticipated, spatial anxiety, statistics anxiety, and mathematics anxiety (AMAS only) are strongly and positively related to one another (r between .49 and .79).

The correlation between statistics anxiety and mathematics anxiety is strong (*r* = .79). It could indicate that the SAS has poor discriminant validity because statistics and mathematics are often confused (e.g., Zeidner, 1991). However, the fact that the respondents were seated in their statistics class when filling the questionnaire suggests that confusion may play a role but cannot be the sole explanation. They may have read the AMAS items while thinking about statistics. This strong correlation can also be explained by a common factor underlying the variables (e.g., gender, general anxiety).

Table 5

Correlations Between Scales' Total Scores

	SAQ	AMAS	SAS-F	LSAS-F	STAI-Y-F	SIMA
SAQ	1					
AMAS	.49 [.43, .54]	1				
SAS-F	.58 [.53, .62]	.79 [.76, .81]	1			
LSAS-F	.53 [.48, .58]	.33 [.27, .39]	.54 [.49, .59]	1		
STAI-Y-F	.41 [.35, .47]	.38 [.32, .44]	.49 [.44, .54]	.61 [.56, .65]	1	
SIMA [a]	.41 [.35, .47]	.79 [.76, .81]	.73 [.69, .76]	.25 [.18, .32]	.36 [.30, .42]	1

Note. Sample size is 776. SAQ is the Spatial Anxiety Questionnaire, AMAS is the Abbreviated Math Anxiety Scale, SAS-F is the French version of the Statistics Anxiety Scale, LSAS-F is the Liebowitz's Social Anxiety Scale, STAI-Y-F is the subscale of the State-Trait Anxiety Inventory and SIMA is the Single-Item Math Anxiety scale; *N* is 778. Number between braces are 95% confidence intervals of the correlation coefficient.

SIMA and AMAS are also strongly correlated (*r* = .79), suggesting that they measure similar constructs of mathematics anxiety, as seen previously. However, SIMA was almost equally correlated to statistics anxiety (*r* = .73). Here again, a confusion between the course contents could be in cause, but a common predictor to both anxieties can also be considered, which is the core hypothesis of the present report. More is discussed in the next section.

Some of the weakest associations concerned trait anxiety. We included trait anxiety to have a baseline level of anxiety. Considering that trait anxiety is the participants' general level of anxiety, it was expected for this variable to be moderately correlated with specific types of anxiety.

Two unexpected correlations were observed, the ones associating social anxiety with statistics anxiety ($r = .54$) and with spatial anxiety ($r = .53$). Considering that statistics anxiety has a social component (anxiety of asking for help), this subscale explains the strong correlation between social anxiety and statistics anxiety, as seen in the next subsection where both LSAS subscales are more strongly correlated with the anxiety of asking for help in statistics. The second association is more difficult to explain. It might be possible that social anxiety (both performance and interaction subscales) requires one or more spatial skill(s), although the present data offers no clue as to why it would be the case. Tarampi et al. (2016) examined the influence of social factors on spatial perspective test. They observed that when the test is framed as a measure of empathy, women's performance is better. Considering the high proportion of women in the current sample, it might explain the association between spatial anxiety and social anxiety. Other recent research has examined spatial perspective taking and social cues (Geer & Ganley, 2023; Gunalp et al., 2019, 2021). In those research, the link between space and social is observed, but with some nuances. Another explanation is that the *Spatial Anxiety Questionnaire* (SAQ) measures a social component. Indeed, some of the items have a social component. For example, an item in the Imagery subscale asks participants if they are anxious when they need to describe in detail the face of a person they met once.

Correlations Between Subscales

The correlation matrix for subscales is presented in Table 7. Within a given domain, the subscales are strongly correlated.

Table 7

Correlations Between Subscales

	SAQ			AMAS		SAS-F			LSAS-F	
	Manipulate	Navigate	Imaginate	Evaluation	Learning	Evaluation	Help	Interp.	Performance	Interaction
Manipulate	1									
Navigate	.43 [.37, .49]	1								
Imagine	.59 [.54, .63]	.45 [.39, .50]	1							
Evaluation	.47 [.42, .53]	.22 [.15, .28]	.29 [.23, .35]	1						
Learning	.45 [.39, .50]	.37 [.31, .43]	.31 [.24, .37]	.62 [.57, .66]	1					
Evaluation	.41 [.35, .47]	.38 [.32, .44]	.28 [.22, .35]	.51 [.46, .56]	.79 [.76, .81]	1				
Help	.36 [.30, .42]	.34 [.27, .40]	.40 [.33, .45]	.50 [.44, .55]	.47 [.41, .52]	.49 [.43, .54]	1			
Interpretation	.59 [.54, .63]	.31 [.24, .37]	.42 [.36, .47]	.73 [.70, .76]	.57 [.52, .62]	.53 [.48, .58]	.52 [.47, .57]	1		
Performance	.37 [.30, .42]	.40 [.34, .46]	.48 [.42, .53]	.30 [.23, .36]	.36 [.30, .42]	.41 [.35, .47]	.56 [.51, .61]	.36 [.30, .42]	1	
Interaction	.33 [.26, .39]	.43 [.37, .49]	.48 [.43, .54]	.22 [.15, .29]	.27 [.20, .33]	.33 [.26, .39]	.53 [.48, .58]	.27 [.20, .34]	.82 [.80, .84]	1

Note. SAQ is the Spatial Anxiety Questionnaire, AMAS is the Abbreviated Math Anxiety Scale, SAS-F is the French version of the Statistics Anxiety Scale, LSAS-F is the Liebowitz's Social Anxiety Scale, STAI-Y-F is the subscale of the State-Trait Anxiety Inventory and SIMA is the Single-Item Math Anxiety scale; *N* is of 778. Number between braces are 95% confidence intervals of the correlation coefficient.

For example, the SAQ correlations for mental manipulation, navigation and imagery varies between .45 and .59. The same occurs for the three SAS subscales (pairwise correlations between .49 and .53). The pairwise correlations are stronger for AMAS (r = .62) and LSAS (r = .82) subscales. The subscales of mathematics anxiety (evaluation and learning of the AMAS) are correlated with the three subscales of statistics anxiety (all r > .47). However, contrary to expectations, anxiety of evaluation in statistics is less correlated to mathematics evaluation anxiety (r = .51) than to mathematics learning anxiety (r = .79). Anxiety of asking for help in statistics is strongly correlated with the social anxiety subscales (r > .53) which is congruent with expectations. Finally, we note the strong correlation between statistics anxiety of interpretation and spatial anxiety to manipulate visual objects (r = .59). We return to this key finding in the last analyses.

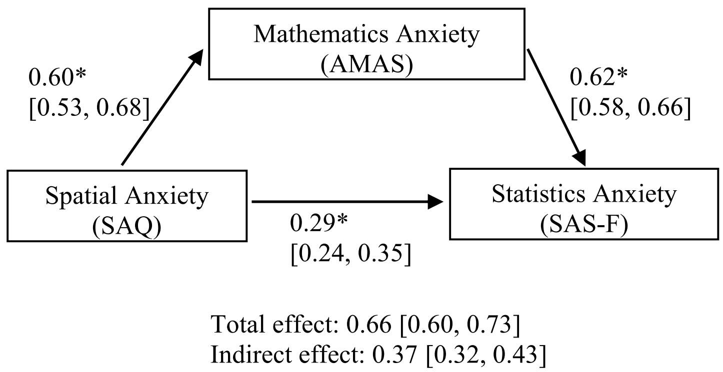
Mediation Analyses

A mediation model was run to determine the unique contribution of spatial anxiety onto statistics anxiety through mathematics anxiety. In this model, anxiety to perform spatial operations would partly explain mathematics anxiety. In turn, mathematics anxiety would partly explain statistics anxiety as this discipline is commonly conceived as a subfield of mathematics. However, this indirect effect would only be partial as spatial anxiety would also directly explain statistics anxiety.

In the first analysis, the results are reported without covariates (as per Simmons et al., 2011). The results, seen in Figure 1, show that the model accounts for 66% of the variance on the total score of the SAS. The indirect effect that goes through AMAS suggests an increase of 0.37 points of anxiety on SAS for each point on SAQ whereas the direct effect indicates about the same result, 0.29. Thus, 56% of the increase in statistics anxiety can be explained by an indirect effect through mathematics anxiety whereas 44% by a direct effect from spatial anxiety.

Figure 1

Mediation Model without Covariates. The Coefficients are the Unstandardized Regression Coefficients.



Note. * denotes cases where the 99.9% confidence interval does not include zero. The 99.9% confidence interval is used here to show that even with a more conservative CI, it still does not include zero.

This result suggests that both spatial anxiety and mathematics anxiety contribute uniquely and account for roughly the same level of statistics anxiety. However, statistics anxiety may most probably be influenced by trait anxiety and to a lesser extent, by social anxiety as well as gender. Consequently, we expanded the model to include these covariates. Even though a gender effect is commonly known to influence anxiety level, it is not clear in the literature if it is a correlate of statistics anxiety (Bui & Alfaro, 2011; DeCesare, 2007; Hsiao & Chiang, 2011; Kesici et al., 2011). Gender differences were examined in more details. Table 8 shows the results of t-tests comparing the mean levels of the types of anxiety. As can be seen, all t-tests are significant as the 95%CI of the mean difference does not include zero.

Table 8

Gender Differences in Mean Scores of Anxiety Types

	Mean difference	95% CI of mean difference	Cohen's d	*t* value and *p*-value
SAQ	−10.3	[−13.8, −6.7]	.59	−5.67 $p < .001$
AMAS	−9.0	[−13.8, −4.1]	.39	−3.62 $p < .001$
SAS	−11.1	[−15.19, −7.0]	.53	−5.34 $p < .001$
LSAS	−7.3	[−11.0, −3.5]	.38	−3.76 $p < .001$
STAI-Y	5.5	[1.6, 9.4]	−.28	2.79 $p = .005$
SIMA	−12.7	[−19.1, −6.2]	.42	−3.88 $p < .001$

Notes. SAQ is the Spatial Anxiety Questionnaire, AMAS is the Abbreviated Math Anxiety Scale, SAS-F is the French version of the Statistics Anxiety Scale, LSAS-F is the Liebowitz's Social Anxiety Scale, STAI-Y-F is the subscale of the State-Trait Anxiety Inventory and SIMA is the Single-Item Math Anxiety scale. The statistics of the AMAS and the SIMA are with unequal variances. Negative scores indicate that women were more anxious than men on average. The *t* and *p* values are presented here for transparency as they are the results of the *t*-tests.

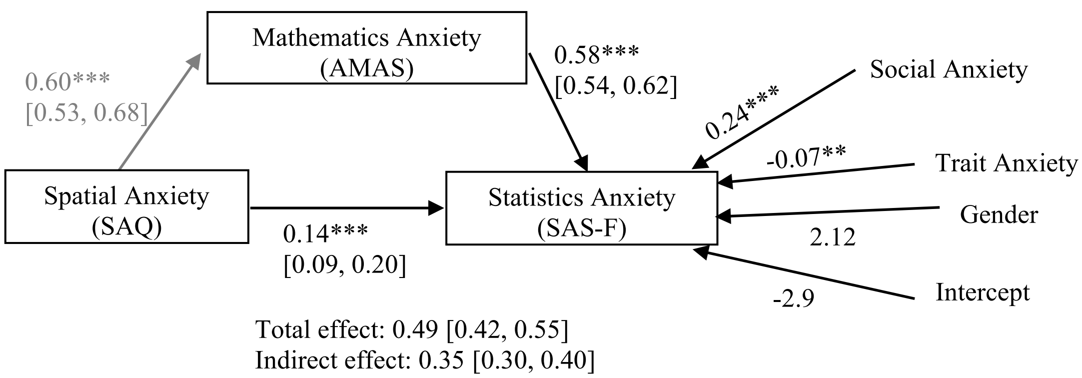
Figure 2 shows the results. Trait anxiety was found to be weakly but significantly associated with statistics anxiety, such that higher levels of trait anxiety were associated with higher levels of statistics anxiety (adding 0.07 points for each point on the STAI-Y; recall that all the scales run from 0 to 100 except for gender, using 0 and 1). Social anxiety contributes to statistics anxiety, where higher levels of social anxiety were associated to higher levels of statistics anxiety (adding 0.24 points for each point of social anxiety). Finally, gender contributes weakly to statistics anxiety ($b = 3.89$, SE = 1.72, $p = .02$).[2] This last result is supported by an additional mediation analysis which corroborates that the gender effect is not specific to statistics

[2] Note that these statistics are presented to ease future meta-analyses as mentioned by a peer-reviewer. However, the 95% CI are included in Figure 2.

anxiety.[3] Instead, it is more strongly characterized as a by-product of the gender effect found in spatial anxiety where gender has an influence more than two times stronger ($b = 9.77$, SE = 1.77, $p < .001$). The gender effect in mathematics anxiety was also similarly explained because mathematics anxiety is weakly predicted by gender when spatial anxiety is considered ($b = 2.43$, SE = 1.99, $p = .22$).

Figure 2

Mediation Model with Covariates. The Coefficients are the Unstandardized Regression Coefficients.



Note. *** denotes cases where the 99.9% confidence interval does not include zero; ** denotes cases where the 99% confidence interval does not include zero; where there is no symbol, the 95% confidence interval includes zero. Gender is coded the following way: men = 1 and women = 2. Social and trait anxieties were put on a range from 0 to 100. The grey path is unchanged from Figure 1.

These analyses show that the covariates (notably social anxiety and trait anxiety) are relevant to understand the relation between spatial anxiety and statistics anxiety. The prominent impact of social anxiety is most plausibly explained by the fact that the dimensions of statistics

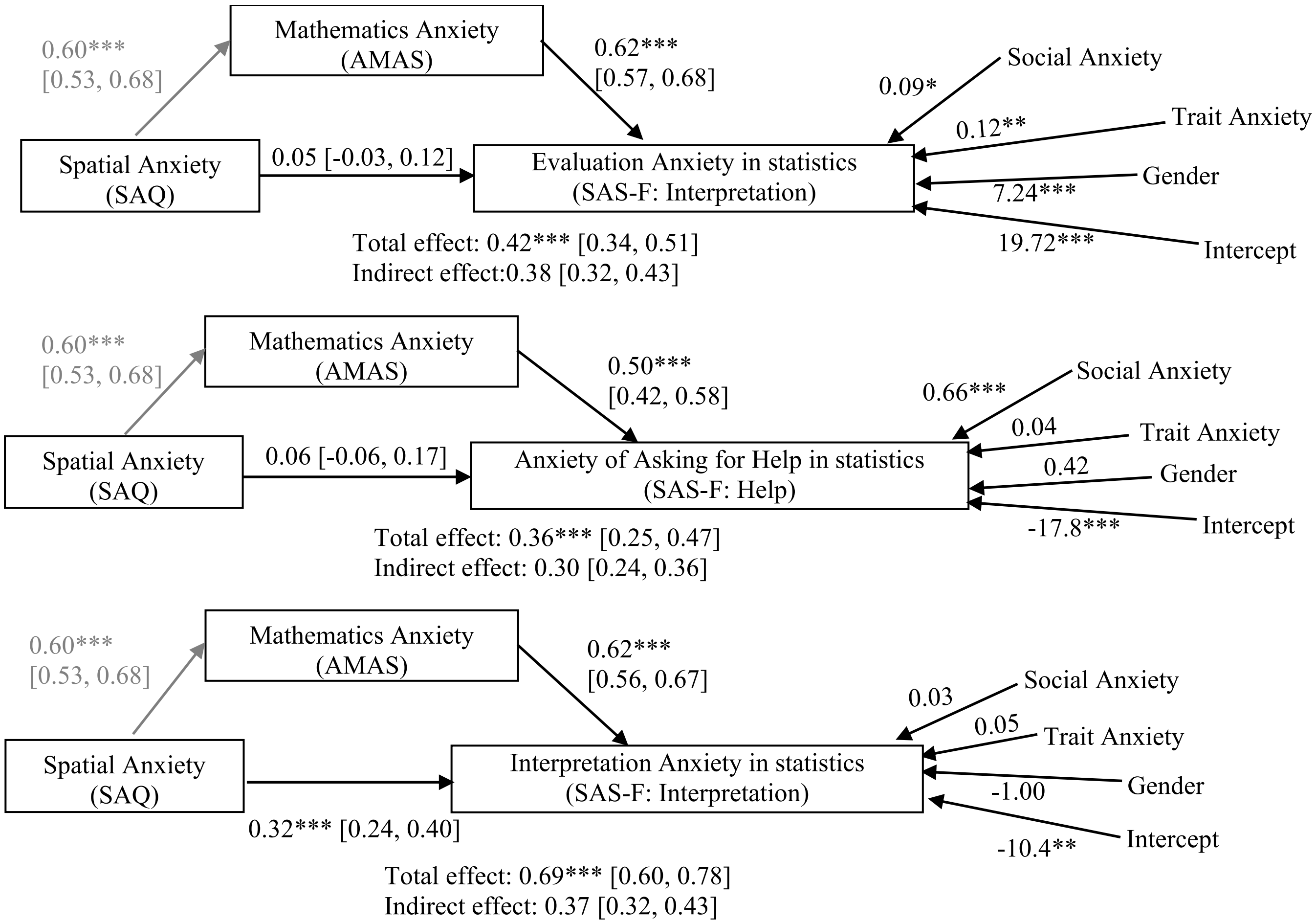
[3] This mediation analysis examines the direct relation of gender on statistics anxiety as well as its indirect relation through spatial anxiety. The numbers reported in the sentence as well as the sentence that follows are from this analysis. We thank the anonymous reviewer who suggested this additional analysis.

anxiety are not affected equally by spatial anxiety. The model was therefore broken down into three models, one for each dimension of statistics anxiety.

Figure 3 shows the three resulting mediation analyses. As seen, with regards to evaluation anxiety in statistics, spatial anxiety is not a contributor. Instead, trait anxiety and gender become the two strongest contributors (b = 0.12, SE = 0.03, p = .001 and b = 7.24, SE = 1.47, p < .001, respectively). Social anxiety is involved in the explanation regarding evaluation anxiety albeit a little less strongly. All three factors explain 57.6% of the variance.

Figure 3

Mediation Model with all Dimensions of Statistics Anxiety. The Coefficients are the Unstandardized Regression Coefficients.



Note. *** denotes cases where the 99.9% confidence interval does not include zero; ** denotes cases where the 99% confidence interval does not include zerc denotes cases where the 95% confidence interval does not include zero; where there is no symbol, the 95% confidence interval includes zero. Gender is coded following way: men = 1 and women = 2. Social and trait anxieties were put on a range from 0 to 100. The grey path is unchanged from Figure 1

By contrast, when examining anxiety of asking for help in statistics, we find that trait anxiety plays no role in explaining it. Instead, social anxiety becomes the most predominant predictor of that statistics anxiety component ($b = 0.66$, SE = 0.06, $p < .001$). Gender and spatial anxiety do not contribute to this component.

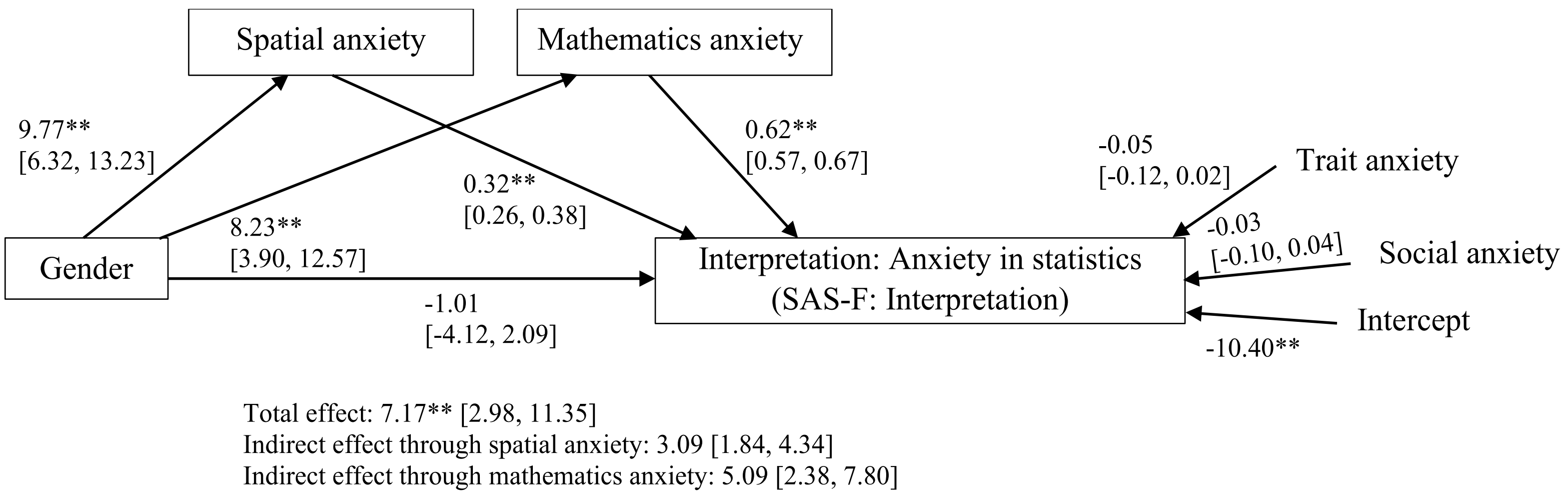
Finally, for anxiety of interpretation in statistics, spatial anxiety becomes a significant contributor whereas social anxiety, trait anxiety and gender are no longer contributors. By comparison, the indirect effect through mathematics anxiety brings 0.37 points of anxiety (0.60×0.62), nearly the same effect as the direct effect (0.32 points).

In sum, in all three subscales, mathematics anxiety remains a strong contributor, suggesting that a portion of statistics anxiety may be explained by mathematics anxiety. Interpretation anxiety seems to be the sole factor originating in spatial anxiety. Evaluation anxiety has strong connections with gender and trait anxiety whereas asking for help anxiety is strongly related to social anxiety.

Because the lack of gender effect regarding anxiety of interpretation is contradicting some past research, this section is expanded with an unplanned mediation analysis focusing on gender with a direct effect on interpretation anxiety in statistics – as this subscale is the most influenced by spatial anxiety – and two effects mediated through spatial anxiety and mathematics anxiety. Figure 4 shows the result.

Figure 4

The Gender Effect Mediated by Spatial Anxiety and Mathematics Anxiety on Interpretation Anxiety in Statistics



*Note.*** denotes cases where the 99% confidence interval does not include zero; where there is no symbol, the 95% confidence interval includes zero. Gender is coded the following way: men = 1 and women = 2. Social and trait anxieties were put on a range from 0 to 100.

As seen, the direct gender effect is negligible whereas its relation through spatial anxiety is important ($b = 9.77 \times 0.32 = 3.13$, SE = 0.64, $p < .001$) and its relation through math anxiety is the largest ($b = 8.23 \times 0.62 = 5.10$, SE = 1.38, $p < .001$).

Given that the direct effect of gender is less than one point added when the participant is a woman (with standard deviation of 23.3 on that subscale, this represents a Cohen's d below 0.04) and given the size of the sample ($n = 778$), it seems safe to conclude that an effect of gender of meaningful magnitude is absent. However, it is necessary to mention that this effect, or lack thereof, can be influenced by where the sample comes from and by the distribution of men and women – i.e., there is a lot more women in psychology than men, which could have inflated the anxiety levels.

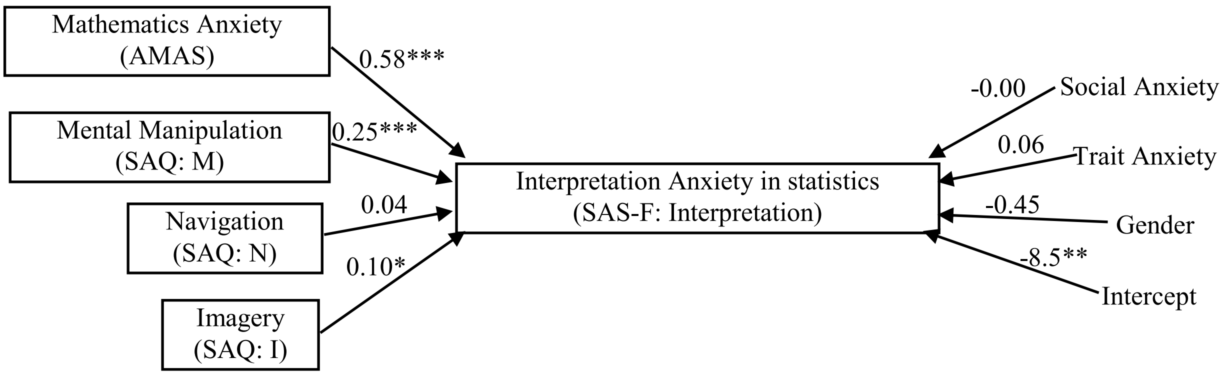
Which Spatial Anxiety Dimension Sustains Interpretation in Statistics?

In this final subsection, we focus more specifically on interpretation anxiety in statistics. This component is possibly the one most related to understanding of statistics concepts and the one most distinct from mathematics (being evaluated is a similar activity in both math and statistics classes; this is also true for asking for help).

To estimate which spatial anxiety component is the most prevalent predictor of interpretation anxiety in statistics, a multiple regression model was estimated, using all three subscales of the SAQ. The goal here was to estimate the contribution of each spatial anxiety component on statistics anxiety. As seen in Figure 5, the most prevalent spatial skills contributor of interpretation anxiety is Mental Manipulation ($b = 0.25$, SE = 0.03, $p < .001$), whereas the least potent is Navigation anxiety ($b = -0.04$, SE = 0.03, $p = .159$). The anxiety to imagine scenes with full details has an intermediate effect ($b = 0.10$, SE = 0.04, $p = .01$).

Figure 5

Multiple Regression Model With All Dimensions of Spatial Anxiety on Interpretation Anxiety in Statistics



Note. *** denotes cases where the 99.9% confidence interval does not include zero; ** denotes cases where the 99% confidence interval does not include zero; * denotes cases where the 95% confidence interval does not include zero ; where there is no symbol, the 95% confidence interval includes zero. Gender is coded the following way: men = 1 and women = 2. Social and trait anxieties were put on a range from 0 to 100.

These results suggest that having statistics anxiety is being anxious about manipulating data and understand the implications of magnitudes expressed with numbers in social and health sciences students. Statistics anxiety is not about being anxious to navigate in space (mentally or physically) as interpretation of data is to provide a static snapshot based on one or a few statistics, not a road map to a final description. Thus, if some limitations underlying spatial anxiety predict statistics discomfort, it might mostly be the apprehension to manipulate visual elements. This is reminiscent of Dehaene et al.'s (1993) SNARC effect, where number magnitudes are mentally represented using a line with smaller magnitudes located spatially on the left (also see Fias, 1996; Dehaene, 2003; Ito & Hatta, 2004; Hubbard et al., 2005; McCrink et al., 2007; Viarouge et al., 2014).

Discussion

This study showed that spatial anxiety, mathematics anxiety, and statistics anxiety are strongly and positively correlated, which suggests communalities – sometimes almost multicollinearity – between those three constructs. The relation between spatial anxiety and statistics anxiety was partially mediated by mathematics anxiety. Through mediation analyses, it was observed that spatial anxiety directly and indirectly influences statistics anxiety. Specifically, mental manipulation anxiety was most closely related to interpretation anxiety in statistics. The results also indicate that the gender effect is less present (maybe even not present at all) in mathematics and statistics anxiety when spatial anxiety is considered. Finally, the analyses indicate that the French translations of the SAQ, and both the AMAS and the SIMA, developed for this study, are reliable measures of spatial anxiety and mathematics anxiety, respectively.

Similarities Between the Anxieties

Moderate-to-strong positive correlations were observed among all the variables. Some are consistent with previous research, for example, the correlation between statistics anxiety and mathematics anxiety ($r = .54$ in Birenbaum & Eylath, 1994; $r = .70$ in Maysick, 1984; and $r = .41$ in Zeidner, 1991, compared to the present $r = .79$). This suggests that the different domains (space, mathematics and statistics) may share a common core and generate similar forms of anxieties. Another explanation is that anxiety in these different domains may be perceived and experienced in similar contexts, when in reality they are distinguishable. Take, for example, the relations between social performance anxiety and anxiety of evaluation in statistics ($r = .36$) and mathematics ($r = .30$). These results indicate that certain components of the different anxiety domains are similar even though the social domain is different from the other two.

To explain this hypothesis, one may look at the opposite direction; the unrelatedness of spatial navigation anxiety and all dimensions of statistics anxiety. Many dimensions of anxiety types are correlated between them, suggesting that they may have common roots. However, some dimensions may be distinguishable in terms of their spatial properties.

For example, spatial navigation anxiety is particularly not related with interpretation anxiety in statistics. Based on Uttal et al.'s (2013) typology, spatial navigation is extrinsic-dynamic, meaning that it involves the relation between moving (or being-moved) objects in a group (e.g., streets names or landmarks in a city). On the other hand, interpreting statistical results could be considered as intrinsic-static and intrinsic-dynamic. Interpretation in statistics involves defining objects in terms of their parts, meaning that one would look at the specification and the relations between the parts of an object (e.g., the different results or numbers of an ANOVA). Also, interpreting results would be static or fixed into place (e.g., a table of correlations) and dynamic or moving (e.g., factor models). Considering that one is extrinsic and the other intrinsic, the unrelatedness between spatial navigation anxiety and interpretation anxiety in statistics makes sense.

It is also possible that spatial, mathematical, and statistical processing share underlying cognitive processes. This supports the claim that people who perform better on measures of spatial processing, also perform better on tests of mathematical ability (Lubinski & Benbow, 1992; McLean & Hitch, 1999; Robinson et al., 1996). This relation is wildly observed (e.g., Gathercole & Pickering, 2000; Guay & McDaniel, 1977; Kyttälä et al., 2003), and is used to predict entry into and success in STEM careers (Casey et al., 1995; Shea et al., 2001; Wai et al., 2009). It has even been argued that "the connection between space and math may be one of the most robust and well-established findings in cognitive psychology" (Mix & Cheng, 2012, p. 198). Furthermore, it was found that numerical and spatial skills influence performance on statistical reasoning tasks (Penna et al., 2014).

Note that the data described above are not longitudinal and, as such, no temporal conclusions can be drawn from the mediation analyses. What can be said is that, in the present data set on social and health sciences students, the relation between spatial anxiety and statistics anxiety can be explained, in part, by mathematics anxiety while the direct effect of spatial anxiety also remains. We could conjecture that initially weaker spatial skills generate early spatial anxiety (observed as early as 6-8 years old; Lauer et al., 2018; Ramirez et al., 2012). This would limit success in mathematics and generate mathematics anxiety (observed as early as approximately 9-10 years old; Harari et al., 2013; Luttenberger et al., 2018). This lack of success set maladaptive emotion regulation strategies (McIntee et al., 2022) mostly attached to anything deemed "mathematics". The literature would benefit from future research emphasizing on longitudinal studies examining this conjectural trajectory on a 15-year span (from 6 years old to college or university where the first statistics classes occur).[4]

Gender Differences

As indicated previously, there is a significant effect of gender on spatial anxiety whereby women report higher anxiety than men. This observation is consistent with similar findings in spatial anxiety (e.g., Casey et al., 1997; Lawton, 1994), and with the findings that adult women also report higher levels of mathematics anxiety (e.g., Maloney & Beilock, 2012). However, the relation between gender and statistics anxiety can be accounted for by spatial anxiety. In other words, in the current sample, gender is not a predictor of statistics anxiety in the mediation

[4] Here, we focus on anxiety as per the scope of the present study. However, note that both anxiety and performance (or underdeveloped skills) are involved.

model. This may indicate that the women's higher level of anxiety in statistics (relative to men's) may be accounted for by a higher level of spatial anxiety. This is congruent with the findings of Delage et al. (2022) and other studies. Wong (2017) noted that the largest gender difference in mathematics involves activities requiring spatial skills (also see Else-Quest et al., 2010; Harnisch et al., 1986). It must be noted that the current sample characteristics may have influenced the results.

The Role of Spatial Tools in the Teaching of Statistics

One message that can be taken out from the present study is that anxiety around spatial processing may play a major role in the level of anxiety experienced by social and health sciences' students in their statistics courses. This raises a question; does the use of graphs and schematics when teaching statistics facilitate the learning of this subject for non-STEM students who are high in statistics anxiety? On one hand, it is possible that the use of graphs and schematics may cause students to experience increased anxiety in their statistics courses, decreasing their ability to learn the concepts. On the other hand, it is also possible that students who are higher in statistics anxiety may present lower spatial skills (indeed, increased spatial anxiety is associated with lower spatial skill; Lyons et al., 2018). Thus, training students in using graphs and schematics may help improve students' skills in visuospatial reasoning, which, in turn, may help them learn to tolerate and eventually lessen the anxiety they experience in class. This could reduce their anxiety towards spatial processing and towards mathematics and statistics. Whether the use of schematics in the teaching of statistics is helpful or harmful for those who are the most anxious about statistics remains an empirical question that will be investigated in future research.

Limitations

A few limitations should be acknowledged. Firstly, in the instructions of each scale, a question is presented (i.e., "are you anxious when..."), and below that question, the items are presented. Before collecting the data, this question was removed from the STAI-Y-F's instructions because it did not match with the type of item. However, this question was removed only in the first block that presented the STAI-Y-F items while assuming it would do the same for the remaining blocks. Because it did not, it could have changed the way participants answered the items (note that these items were randomly placed into three blocks). Also, the fit indices of the invariance analysis were below recommendations. This may indicate that both continents have poor-to-acceptable invariance. As such, the conclusions need to be taken carefully.

Another limitation in the present study is the fact that all types of anxiety were measured by questionnaires, which can generate social desirability bias in participants. Participants under-report their thoughts, emotions, and behaviour that are considered inappropriate by society and overstate those that are considered desirable (Donaldson & Grant-Vallone, 2002). Also, the questionnaires were presented online, which may have activated participants' spatial anxiety before answering the items. If the measures were done paper-pencil, participants may not have reported such high spatial anxiety.

Furthermore, only one type of validity was tested with the present sample. Indeed, factor analyses only test the instruments' construct validity. However, there are other types of validity that were not examined here. It is possible that the instruments' construct validity is good, but that their convergent and divergent validity are not. This would have influenced the results by providing responses unrelated to the constructs we were trying to measure. Take the correlation between the AMAS and the SAS (*r* = .79) for example. This result may indicate poor divergent validity. However, the opposite argument can be made as they only have 55% of shared variance. If further investigations of the measures' convergent and divergent validity have been done, this argumentation could have been solved. In the results section, it is argued that a correlation of .30 between two variables represents a relation based on the participants' baseline level of anxiety. However, this has not been formally tested. A regression or a median split analysis should have been done to provide statistical support. It is also possible that the sample's characteristics (underrepresentation of males and overrepresentation of females in particular) influenced the results of the present study. While it is common to have a gender disparity in social sciences, research similar to this one indicates that the discipline can be evenly distributed across gender in studies (e.g., Geer & Ganley, 2023; their sample was composed of 134 men and 144 women undergraduate students recruited from the research pool of the authors' psychologie department).

Participants were all enrolled in a statistics course at the time of the study, which could have exacerbated their anxiety towards the subject, influencing the results. With these potentially exaggerated anxiety levels, the results presented here may appear stronger or more positive. These results may have tainted the interpretations in favor of the hypotheses proposed herein. Students in programs similar to psychology may do so because of their high mathematics or spatial anxiety or because they are avoiding STEM majors. Furthermore, data was collected at one time point, limiting the interpretations to correlational associations. Future research is needed to investigate potential causal links between the types of anxiety. Lastly, the shared variance between the variables could partly be explained by the common data collection strategy used to measure them (i.e., all questionnaires embedded in a single survey).

Conclusion

To our knowledge, this study was the first to examine the possibility of a relation between spatial anxiety and statistics anxiety, building a bridge between two research domains (i.e., spatial representations and statistics education). The present study is a step towards the ultimate goal of improving the education of statistics. Also, the validation of the French version of three scales (i.e., SAQ-F, AMAS-F and SIMA-F) could help future research on spatial anxiety and mathematics anxiety in French-speaking populations. Statistics courses are difficult to teach, and the educational techniques used to facilitate the students' learning do not solve the origin of the problem (Cousineau & Harding, 2017). Even though research on statistics anxiety progresses rapidly, more studies are needed to truly understand this type of anxiety and eventually reduce it.

Funding

This research was supported by a fellowship from the University of Ottawa to the first author, by a research grant from the Faculty of Social Sciences of the University of Ottawa and by funding from the Fondation Maison des Sciences de l'Homme to the last author.

Acknowledgments

We would like to thank Jean-Christophe Goulet-Pelletier and Marc-André Goulet for their comments on an earlier version of this text. Competing Interests: The authors have declared that no competing interests exist.

Data Availability

For this article, a data set is freely available (Gibeau et al., 2022).

Supplementary Materials

The Supplementary Materials contain the following items (for access see Index of Supplementary Materials below): Table S1 shows the descriptive statistics of all scales divided by continent; Figure S1 is the visual support for the confirmatory factor analysis of the SAQ-F; Figure S2 is the visual support for the confirmatory factor analysis of the AMAS-F

Index of Supplementary Materials

Gibeau, R., Maloney, E. A., Béland, S., Lalande, D., Cantinotti, M., Williot, A., Chanquoy, L., Simon, J., Boislard-Pépin, M., & Cousineau, D. (2022). Supplementary materials to "The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender" [Research data and materials]. OSF. https://osf.io/jpv5n/

References

Alvarez-Vargas, D., Abad, C., & Pruden, S. M. (2020). Spatial anxiety mediates the sex difference in adult mental rotation test performance. *Cognitive Research: Principles and Implications, 5*, Article 31. https://doi.org/10.1186/s41235-020-00231-8

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181-185. https://doi.org/10.1111/1467-8721.00196

Atit, K., Power, J. R., Pigott, T., Lee, J., Geer, E. A., Uttal, D. H., Ganley, C. M., & Sorby, S. A. (2022). Examining the relations between spatial skills and mathematical performance: A meta-analysis. *Psychonomic Bulletin & Review, 29*(3), 699-720. https://doi.org/10.3758/s13423-021-02012-w

Atit, K., & Rocha, K. (2020). Examining the relations between spatial skills, spatial anxiety, and K-12 teacher practice. *Mind, Brain, and Education, 15*(1), 139-148. https://doi.org/10.1111/mbe.12274

Atit, K., Uttal, D. H., & Stieff, M. (2020). Situating space: Using a discipline-focused lens to examine spatial thinking skills. *Cognitive Research: Principles and Implications, 5*, Article 19. https://doi.org/10.1186/s41235-020-00210-z

Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *Journal of Creative Behavior, 42*(2), 75-105. https://doi.org/10.1002/j.2162-6057.2008.tb01289.x

Baloğlu, M. (2003). Individual differences in statistics anxiety among college students. *Personality and Individual Differences, 34*(5), 855-865. https://doi.org/10.1016/S0191-8869(02)00076-4

Baloğlu, M., Deniz, M. E., & Kesici, Ş. (2011). A descriptive study of individual and cross-cultural differences in statistics anxiety. *Learning and Individual Differences, 21*(4), 387-391. https://doi.org/10.1016/j.lindif.2011.03.003

Banerjee, P. A. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education, 3*(1), Article 1178441. https://doi.org/10.1080/2331186X.2016.1178441

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin, 147*(2), 134-168. https://doi.org/10.1037/bul0000307

Beard, C., Rodriguez, B. F., Moitra, E., Sibrava, N. J., Bjornsson, A., Weisberg, R. B., & Keller, M. B. (2011). Psychometric properties of the Liebowitz Social Anxiety Scale (LSAS) in a longitudinal study of African Americans with anxiety disorders. *Journal of Anxiety Disorders, 25*(5), 722-726. https://doi.org/10.1016/j.janxdis.2011.03.009

Béland, S., Cousineau, D., & Loye, N. (2016). Les dix commandements du nouvel homo statisticus. *McGill Journal of Education / Revue des sciences de l'éducation de McGill, 51*(2), 947-960. https://doi.org/10.7202/1038612ar

Béland, S., Cousineau, D., & Loye, N. (2017). Utiliser le coefficient omega de McDonald à la place de l'alpha de Cronbach. *McGill Journal of Education / Revue des sciences de l'éducation de McGill, 52*(3), 791-804. https://doi.org/10.7202/1050915ar

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *The Journal of Experimental Education, 57*(3), 247-261. https://doi.org/10.1080/00220973.1989.10806509

Berman, N. C., Wheaton, M. G., McGrath, P., & Abramowitz, J. S. (2010). Predicting anxiety: The role of experiential avoidance and anxiety sensitivity. *Journal of Anxiety Disorders,* *24*(1). 109-113. https://doi.org/10.1016/j.janxdis.2009.09.005

Birenbaum, M., & Eylath, S. (1994). Who is afraid of statistics? Correlates of statistics anxiety among students of educational sciences. *Educational Research, 36*(1), 93-98. https://doi.org/10.1080/0013188940360110

Bourne, V. J. (2018). Exploring statistics anxiety: Contrasting mathematical, academic performance and trait psychological predictors. *Psychology Teaching Review, 24*(1), 35-43.

Brebner, J. (2003). Gender and emotions. *Personality and Individual Differences, 34*(3), 387-394. https://doi.org/10.1016/S0191-8869(02)00059-4

Bui, N. H., & Alfaro, M. A. (2011). Statistics anxiety and science attitudes: Age, gender, and ethnicity factors. *College Student Journal, 45*(3), 573-585.

Cantinotti, M., Lalande, D., Ferlatte, M.-A., & Cousineau, D. (2017). Validation de la version francophone du Questionnaire d'anxiété statistique (SAS-F-24) [Validation of the French version of the Statistics Anxiety Scale (SAS-F)]. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 49*(2), 133-142. https://doi.org/10.1037/cbs0000074

Cappelleri, J., Zou, K., Bushmakin, A. G., Alvir, J. M. J., Alemayehu, D., & Symonds, T. (2014). *Patient-reported outcomes: Measurement, implementation and interpretation*. Boca Raton, FL, USA: CRC Press.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33*(4), 669-680. https://doi.org/10.1037/0012-1649.33.4.669

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31*(4), 697-705. https://doi.org/10.1037/0012-1649.31.4.697

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*(1), 2-11. https://doi.org/10.1080/15248372.2012.725186

Chew, P. K. H., Dillon, D. B., & Swinbourne A. L. (2018) An examination of the internal consistency and structure of the Statistical Anxiety Rating Scale (STARS). *PLoS ONE, 13*(3), Article e0194195. https://doi.org/10.1371/journal.pone.0194195

Comfrey, A. L., & Lee, H. B. (1992). A first course in factor analysis (2nd ed.). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Cousineau, D. (2020). How many decimals? Rounding descriptive and inferential statistics based on measurement precision. *Journal of Mathematical Psychology, 97*, Article 102362. https://doi.org/10.1016/j.jmp.2020.102362

Cousineau, D., & Harding, B. (2017). Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions. *Revue de psychoéducation, 46*(2), 397-419. https://doi.org/10.7202/1042257ar

Cruise, R., Cash, R., & Bolton, D. (1985). Development and validation of an instrument to measure statistical anxiety. In *Annual meeting of the Statistical Education Section* (pp. 92-97). American Statistical Association.

Cui, S., Zhang, J., Guan, D., Zhao, X., & Si, J. (2019). Antecedents of statistics anxiety: An integrated account. *Personality and Individual Differences, 144*(2019), 79-87. https://doi.org/10.1016/j.paid.2019.02.036

DeCesare, M. (2007). "Statistics anxiety" among sociology majors: A first diagnosis and some treatment options. *Teaching Sociology, 35*(4), 360-367. https://doi.org/10.1177/0092055X0703500405

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7*(4), 145-147. https://doi.org/10.1016/S1364-6613(03)00055-X

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396. https://doi.org/10.1037/0096-3445.122.3.371

Delage, V., Trudel, G., Retanal, F., & Maloney, E. A. (2022). Spatial anxiety and spatial ability: Mediators of gender differences in math anxiety. *Journal of Experimental Psychology: General, 151*(4), 921-933. https://doi.org/10.1037/xge0000884

DeMaria Mitton, P. A. (1987). *Locus-of-control, gender and type of major as correlates to statistics anxiety in college* (Publication No. 8721001) [Doctoral dissertation, The American University]. ProQuest Dissertations Publishing.

DeVaney, T. A. (2010). Anxiety and attitude of graduate students in on-campus vs. online statistics courses. *Journal of Statistics Education, 18*(1), 1-15. https://doi.org/10.1080/10691898.2010.11889472

de Hevia, M. D., Vallar, G., & Girelli, L. (2008). Visualizing numbers in the mind's eye: The role of visuo-spatial processes in numerical abilities. *Neuroscience and Biobehavioral Reviews, 32*(8), 1361-1372. https://doi.org/10.1016/j.neubiorev.2008.05.015

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology, 17*(2), 245-260.

https://doi.org/10.1023/A:1019637632584

Douglas, H. P., & LeFevre, J.-A. (2018). Exploring the influence of basic cognitive skills on the relation between math performance and math anxiety. *Journal of Numerical Cognition, 3*(3), 642-666. https://doi.org/10.5964/jnc.v3i3.113

Eduljee, N. B., & LeBourdais, P. (2015). Gender differences in statistics anxiety with undergraduate college students. *The International Journal of Indian Psychology, 2*(3), 69-82. https://doi.org/10.25215/0203.028

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127. https://doi.org/10.1037/a0018053

Eme, P.-E., & Marquer, J. (1999). Individual strategies in a spatial task and how they relate to aptitudes. *European Journal of Psychology of Education, 14*(1), 89-108. https://doi.org/10.1007/BF03173113

Ferguson, A., Maloney, E. A., Fugelsang, J., & Risko, E. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences, 39*, 1-12. https://doi.org/10.1016/j.lindif.2015.02.007

Fias, W. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC Effect. *Mathematical Cognition, 2*(1), 95-110. https://doi.org/10.1080/135467996387552

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics—A “hopeless” issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22*(4), 497-514. https://doi.org/10.1007/BF03173468

Fullerton, J., & Umphrey, D. (2016). Statistics anxiety and math aversion among advertising students. *Journal of Advertising Education, 20*(1-2), 135-143. https://doi.org/10.1177/10980482160201-216

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences* (10th anniversary ed.). Basic Books.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*(2), 177-194. https://doi.org/10.1348/000709900158047

Gauthier, J., & Bouchard, S. (1993). Adaptation canadienne-française de la forme révisée du State–Trait Anxiety Inventory de Spielberger [A French-Canadian adaptation of the revised version of Spielberger's State–Trait Anxiety Inventory]. *Canadian Journal of Behavioural Science, 25*(4), 559-578. https://doi.org/10.1037/h0078881

Geer, E. A., & Ganley, C. (2023). Sex differences in social and spatial perspective taking: A replication and extension of Tarampi et al. (2016). *Quarterly Journal of Experimental Psychology, 76*(1), 93-108. https://doi.org/10.1177/17470218221085117

Geer, E., A, Barroso, C., Conlon, R. A., Dasher, J. M., & Ganley, C. M. (2024). A meta-analytic review of the relation between spatial anxiety and spatial skills. *Psychological Bulletin, 150*(4), 464-486. https://psycnet.apa.org/doi/10.1037/bul0000420

Geer, E. A., Quinn, J. M., & Ganley, C. M. (2019). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology, 55*(3), 637-652. https://doi.org/10.1037/dev0000649

Goulet-Pelletier, J. C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's d family. *The Quantitative Methods for Psychology, 14*(4), 242-265. https://doi.org/10.20982/tqmp.14.4.p242

Guay, R. B., & McDaniel, E. D. (1977). The relationship between mathematics achievement and spatial abilities among elementary school children. *Journal for Research in Mathematics Education, 8*(3), 211-215. https://doi.org/10.2307/748522

Gunalp, P., Chrastil, E. R., & Hegarty, M. (2021). Directionality eclipses agency: How both directional and social cues improve spatial perspective taking. *Psychonomic Bulletin & Review, 28*, 1289-1300. https://doi.org/10.3758/s13423-021-01896-y

Gunalp, P., Moossaian, T., & Hegarty, M. (2019). Spatial perspective taking: Effects of social, directional, and interactive cues. *Memory & Cognition, 47*, 1031-1043. https://doi.org/10.3758/s13421-019-00910-y

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2013). Teachers' spatial anxiety relates to 1st- and 2nd-graders' spatial learning. *Mind, Brain, and Education, 7*(3), 196-199. https://doi.org/10.1111/mbe.12027

Hambrick, D. Z., & Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science, 20*(5), 275-279. https://doi.org/10.1177/0963721411422061

Hancock, G. R., & An, J. (2020). A closed-form alternative for estimating omega reliability under unidimensionality. *Measurement: Interdisciplinary Research and Perspectives, 18*(1), 1-14. https://doi.org/10.1080/15366367.2019.1656049

Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics anxiety in young children: An exploratory study. *The Journal of Experimental Education, 81*(4), 538-555. https://doi.org/10.1080/00220973.2012.727888

Harding, B., Tremblay, C., & Cousineau, D. (2014) Standard errors: A review and evaluation of standard error estimators using Monte Carlo simulations. *The Quantitative Methods for Psychology, 10*(2), 107-123. https://doi.org/10.20982/tqmp.10.2.p107

Harnisch, D. L., Steinkamp, M. W., Tsai, S.-L., & Walberg, H. J. (1986). Cross-national differences in mathematics attitude and achievement among seventeen-year-olds. *International Journal of Educational Development, 6*(4), 233-244. https://doi.org/10.1016/0738-0593(86)90027-1

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for quantifying reliability. But... *Communication Methods and Measures, 14*(1), 1-24. https://doi.org/10.1080/19312458.2020.1718629

Heeren, A., Maurage, P., Rossignol, M., Vanhaelen, M., Peschard, V., Eeckhout, C., & Philippot, P. (2012). Self-report version of the Liebowitz Social Anxiety Scale: Psychometric properties of the French Version. *Canadian Journal of Behavioural Science, 44*(2), 99-107. https://doi.org/10.1037/a0026249

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., & Liebowitz, M. R. (1999). Psychometric properties of the Liebowitz Social Anxiety Scale. *Psychological Medicine, 29*(1), 199-212. https://doi.org/10.1017/S0033291798007879

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment, 10*(2), 178-182. https://doi.org/10.1177/1073191103010002008

Hsiao, T.-y., & Chiang, S. (2011). Gender differences in statistics anxiety among graduate students learning English as a foreign language. *Social Behavior and Personality, 39*(1), 41-42. https://doi.org/10.2224/sbp.2011.39.1.41

Hubbard, E., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience, 6*, 435-448. https://doi.org/10.1038/nrn1684

Hund, A., & Minarik, J. (2006). Getting from here to there: Spatial anxiety, wayfinding strategies, direction type, and wayfinding efficiency. *Spatial Cognition & Computation, 6*(3), 179-201. https://doi.org/10.1207/s15427633scc0603_1

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32*(4), 662-673. https://doi.org/10.3758/BF03195857

Jordan, K., Wüstenberg, T., Heinze, H., Peters, M., & Jäncke, L. (2002) Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia, 40*(13), 2397-2408. https://doi.org/10.1016/S0028-3932(02)00076-3

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kesici, Ş., Baloğlu, M., & Deniz, M. E. (2011). A descriptive study of individual and cross-cultural differences in statistics anxiety. *Learning and Individual Differences, 21*(4), 387-391. https://doi.org/10.1016/j.lindif.2011.03.003

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education, 3*(1), 1-9. https://doi.org/10.1080/10691898.1995.11910479

Knowles, K. A., & Olatunji, B. O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the State-Trait Anxiety Inventory. *Clinical Psychology Review, 82*, Article 101928. https://doi.org/10.1016/j.cpr.2020.101928

Kremmyda, O., Hüfner, K., Flanagin, V., Hamilton, D., Linn, J., Strupp, M., Jahn, K., & Brandt, T. (2016). Beyond dizziness: Virtual navigation, spatial anxiety and hippocampal volume in bilateral vestibulopathy. *Frontiers in Human Neuroscience, 10*, Article 139. https://doi.org/10.3389/fnhum.2016.00139

Kyttälä, M., Aunio, P., Lehto, J. E., Van Luit, J., & Hautamäki, J. (2003). Visuospatial working memory and early numeracy. *Educational and Child Psychology, 20*(3), 65-76.

Lalande, D., Cantinotti, M., Williot, A., Gagnon, J., & Cousineau, D. (2019). Three pathways from achievement goals to academic performance in an undergraduate statistics course. *Statistics Education Research Journal, 18*(1), 94-105. https://doi.org/10.52041/serj.v18i1.152

Lauer, J. E., Esposito, A. G., & Bauer, P. J. (2018). Domain-specific anxiety relates to children's math and spatial performance. *Developmental Psychology, 54*(11), 2126-2138. https://doi.org/10.1037/dev0000605

Lawton, C. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles, 30*(11-12), 765-779. https://doi.org/10.1007/BF01544230

Lemieux, C. L., Collins, C. A., & Watier, N. N. (2019). Gender differences in metacognitive judgments and performance on a goal-directed wayfinding task. *Journal of Cognitive Psychology, 31*(4), 453-466. https://doi.org/10.1080/20445911.2019.1625905

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports, 22*(4), 335-338. https://doi.org/10.1080/00036810500317730

Lubinski, D., & Benbow, C. P. (1992). Gender differences in abilities and preferences among the gifted: Implications for the math-science pipeline. *Current Directions in Psychological Science, 1*(2), 61-66. https://doi.org/10.1111/1467-8721.ep11509746

Luttenberger, S., Wimmer, S., & Paechter, M. (2018). Spotlight on math anxiety. *Psychology*

Research and Behavior Management, 11, 311-322. https://doi.org/10.2147/PRBM.S141421

Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition, 4*(3), 526-553. https://doi.org/10.5964/jnc.v4i3.154

Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European Journal of Psychology of Education, 27*(4), 483-498. https://doi.org/10.1007/s10212-011-0090-5

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*(1), 593-614. https://doi.org/10.1146/annurev.psych.58.110405.085542

Mackinnon, S. P., Alexander, S. M., Chen, R., Cribbie, R. A., Flett, G. L., & Hill, T. G. (2025). Perfectionism, anxiety sensitivity, and statistics anxiety: A test of the vulnerability-stress model using a 2-wave longitudinal study. *Journal of Research in Personality, 116*, Article 104611. https://doi.org/10.1016/j.jrp.2025.104611

Malanchini, M., Rimfeld, K., Shakeshaft, N., Rodic, M., Schofield, K., Selzam, S., Dale, P., Petrill, S., & Kovas, Y. (2017). The genetic and environmental aetiology of spatial, mathematics and general anxiety. *Scientific Reports, 7*, Article 42218. https://doi.org/10.1038/srep42218

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences, 16*(8), 404-406. https://doi.org/10.1016/j.tics.2012.06.008

Maloney, E. A., Waechter, S., Risko, E. F., & Fugelsang, J. A. (2012). Reducing the sex difference in math anxiety: The role of spatial processing ability. *Learning and Individual Differences, 22*(3), 380-384. https://doi.org/10.1016/j.lindif.2012.01.001

Mandap, C. M. (2016). Examining gender differences in statistics anxiety among college students. *International Journal of Education and Research, 4*(6), 358-366. https://www.ijern.com/journal/2016/June-2016/31.pdf

Martinez, J. G. R. (1987). Preventing math anxiety: A prescription. *Academic Therapy, 23*(2), 117-125. https://doi.org/10.1177/105345128702300201

Maysick, D. J. (1984). *The relationship of self-concept, sex, age, major, anxiety, Mathematics background, and mathematics ability to the level of mathematics and statistical anxiety among college juniors* (Publication No. 568) [Doctoral dissertation, Andrews University]. James White Library.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Perception & Psychophysics, 69*(8), 1324-1333. https://doi.org/10.3758/BF03192949

McIntee, S.-E., Goulet-Pelletier, J.-C., Williot, A., Deck-Léger, E., Lalande, D., Cantinotti, M., & Cousineau, D. (2022). (Mal)adaptive cognitions as predictors of statistics anxiety. *Statistics Education Research Journal, 21*, Article 5. https://doi.org/10.52041/serj.v21i1.364

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74*(3), 240-260. https://doi.org/10.1006/jecp.1999.2516

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543. https://doi.org/10.1007/BF02294825

Mix, K. S., & Cheng, Y-L. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior, 42*, 197-243. https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General, 145*(9), 1206-1227. https://doi.org/10.1037/xge0000182

Newcombe, N., & Shipley, T. (2015). *Thinking about spatial thinking: New typology, new assessments*. In J. S. Gero (Ed.), Studying visual and spatial reasoning for design creativity (pp. 179-192). Springer. https://doi.org/10.1007/978-94-017-9297-4_10

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The Single-Item Math Anxiety Scale: An alternative way of measuring mathematical anxiety. *Journal of Psychoeducational Assessment, 32*(4), 306-317. https://doi.org/10.1177/0734282913508528

Odic, D., & Wojcik, E. H. (2020). The publication gender gap in psychology. *American Psychologist, 75*(1), 92-103. https://doi.org/10.1037/amp0000480

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education, 29*(1), 3-19. https://doi.org/10.1080/0260293042000160384

Onwuegbuzie, A., DaRos, D., & Ryan, J. (1997). The components of statistics anxiety: A phenomenological study. *Focus on Learning Problems in Mathematics, 19*(4), 11-35.

Onwuegbuzie, A., & Wilson, V. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments–A comprehensive review of the literature. *Teaching in Higher Education,* *8*(2), 195–209. https://doi.org/10.1080/1356251032000052447

Ouyang, X., Zhang, X., & Zhang, Q. (2022). Spatial skills and number skills in preschool children: The moderating role of spatial anxiety. *Cognition, 225*, Article 105165. https://doi.org/10.1016/j.cognition.2022.105165

Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., & Papousek, I. (2017). Mathematics anxiety and statistics anxiety: Shared but also unshared components and antagonistic contributions to performance in statistics. *Frontiers in Psychology, 8*, Article 1196. https://doi.org/10.3389/fpsyg.2017.01196

Penna, M. P., Agus, M., Peró-Cebollero, M., Guàrdia-Olmos, J., & Pessa, E. (2014). The use of imagery in statistical reasoning by university undergraduate students: A preliminary study. *Quality & Quantity, 48*, 173-187. https://doi.org/10.1007/s11135-012-9757-5

Primi, C., Busdraghi, C., Tomasetto, C., Morsanyi, K., & Chiesi, F. (2014). Measuring math anxiety in Italian college and high school students: Validity, reliability and gender invariance of the Abbreviated Math Anxiety Scale (AMAS). *Learning and Individual Differences, 34*, 51-56. http://dx.doi.org/10.1016/j.lindif.2014.05.012

Primi, C., & Chiesi, F. (2018). The role of mathematics anxiety and statistics anxiety in Learning Statistics. In *Tenth International Conference on Teaching Statistics*. https://flore.unifi.it/retrieve/handle/2158/1142287/377210/ICOTS10_5E2.pdf

Ramirez, G., Gunderson, E., Levine, S., & Beilock, S. (2012). Spatial anxiety relates to spatial abilities as a function of working memory in children. *The Quarterly Journal of Experimental Psychology, 65*(3), 474-487. https://doi.org/10.1080/17470218.2011.616214

Ratliff, K. R., Saunders, J., & Levine, S. C. (2009, November). *The cognitive basis of individual* *differences in children's mental transformation skill* [Poster presentation]. Annual meeting of the Psychonomics Society, Boston, MA, USA.

Robinson, N. M., Abbott, R. D., Berninger, V. W., & Busse, J. (1996). Structure of abilities in math precocious young children: Gender similarities and differences. *Journal of Educational Psychology, 88*(2), 341-352. https://doi.org/10.1037/0022-0663.88.2.341

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rotzer, S., Loennecker, T., Kucien, K., Martin, E., Klaver, P., & van Aster, M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. *Neuropsychologia, 47*(13), 2859-2865. https://doi.org/10.1016/j.neuropsychologia.2009.06.009

Schmitz, S. (1999). Gender differences in acquisition of environmental knowledge related to wayfinding behavior, spatial anxiety and self-estimated environmental competencies. *Sex Roles, 41*, 71-93. https://doi.org/10.1023/A:1018837808724

Shaughnessy, M. J. (2010, August). *Statistics for all—The flip side of quantitative reasoning. National Council of Teachers of Mathematics*. https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/J_-Michael-Shaughnessy/Statistics-forAll%E2%80%94the-Flip-Side-of-Quantitative-Reasoning

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*(3), 604-614. https://doi.org/10.1037/0022-0663.93.3.604

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant.

Psychological Science, 22, 1359-1366. https://doi.org/10.1177/0956797611417632

Sokolowski, H. M., Hawes, Z., & Lyons, I. (2019) What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition, 182*, 193-212. https://doi.org/10.1016/j.cognition.2018.10.005

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Tolvanen, A., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences, 69*, 173-181. https://doi.org/10.1016/j.lindif.2018.12.005

Spielberger, C. D., Gonzalez-Reigosa, F., & Martinez-Urrutia, A. (1971). Development of the Spanish edition of the State-Trait Anxiety Inventory. *Interamerican Journal of Psychology, 5*, 145-158.

Stieff, M. (2004). *A localized model of spatial cognition in chemistry* (Publication No. 3132612) [Doctoral dissertation, Northwestern University]. ProQuest.

Tarampi, M. R., Heydari, N., & Hegarty, M. (2016). A tale of two types of perspective taking: Sex differences in spatial ability. *Psychological Science, 27*(11), 1507-1516. https://doi.org/10.1177/0956797616667459

Tosto, M. G., Hanscombe, K. B., Haworth, C. M. A., Davis, O. S. P., Petrill, S. A., Dale, P. S., Malykh, S., Plomin, R., & Kovas, Y. (2014). Why do spatial abilities predict mathematical performance? *Developmental Science, 17*(3), 462-470. https://doi.org/10.1111/desc.12138

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? *Psychology of Learning and Motivation, 57*, 147-181. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science, 22*(5), 367-373. https://doi.org/10.1177/0963721413484756

Vasilyeva, M., & Lourenco, S. F. (2010). *Spatial development*. In W. F. Overton & R. M. Lerner (Eds.), The handbook of life-span development: Vol. 1. Cognition, biology, and methods (pp. 720-753). John Wiley & Sons. https://doi.org/10.1002/9780470880166.hlsd001020

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). I. Spatial skills, their development, and their links to mathematics. *Monographs of the Society for Research in Child Development, 82*(1), 7-30. https://doi.org/10.1111/mono.12280

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology, 126*, 37-51. https://doi.org/10.1016/j.jecp.2014.02.012

Viarouge, A., Hubbard, E. M., & Dehaene, S. (2014). The organization of spatial reference frames involved in the SNARC effect. *Quarterly Journal of Experimental Psychology, 67*(8), 1484-1499. https://doi.org/10.1080/17470218.2014.897358

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicothema, 20*(1), 174-180. https://reunido.uniovi.es/index.php/PST/article/view/8638

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817-835. https://doi.org/10.1037/a0016127

Walsh, J. J., & Ugumba-Agwunobi, G. (2002). Individual differences in statistics anxiety: The roles of perfectionism, procrastination and trait anxiety. *Personality and Individual Differences, 33*(2), 239-251. https://doi.org/10.1016/S0191-8869(01)00148-9

Wong, W. I. (2017). The space-math link in preschool boys and girls: Importance of mental transformation, targeting accuracy, and spatial anxiety. *British Journal of Developmental Psychology, 35*(2), 249-266. https://doi.org/10.1111/bjdp.12161

Xu, K. (2012). *Multiple group measurement: Invariance analysis in Lavaan* [PowerPoint slides]. https://users.ugent.be/~yrosseel/lavaan/multiplegroup6Dec2012.pdf

Yao, S.-N., Note, I., Fanget, F., Albuisson, E., Bouvard, M., Jalenques, I., & Cottraux, J. (1999). L'anxiété sociale chez les phobiques sociaux: Valiation de l'échelle d'anxiété sociale de Liebowitz (version française) [Social anxiety in social phobics: Validation of Liebowitz's social anxiety scale–French version]. *L'Encéphale: Revue de psychiatrie clinique biologique et thérapeutique, 25*(5), 429-435.

Zanakis, S., & Valenzi, E. (1997). Student anxiety and attitudes in business statistics. *Journal of Education for Business, 73*(1), 10-16. https://doi.org/10.1080/08832329709601608

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology, 61*(3), 319-328. https://doi.org/10.1111/j.2044-8279.1991.tb00989.x

Chapter 3: Study B

The Relation of Spatial Skills, Spatial Memory Span and two Anxiety Types With Statistics Anxiety in European and North American University Students

Reference

Gibeau, R.-M., Béland, S., Maloney, E. A., Cantinotti, M., Chanquoy, L., Lalande, D., Simon, J., Paulis, C., Plancher, G., Boislard, M.-A. & Cousineau, D. (in press). The relation of spatial skills and spatial memory span with statistics anxiety. *British Journal of Educational Psychology*. DOI: 10.1111/bjep.70063

Contribution

Study conceptualisation, coordination of ethical requests, data collection, data analysis and redaction/revision of the manuscript.

Highlights

- Uses self-reported scores but also actual scores obtained with experimental task.
- Considers both resources (working memory) and skills (performance).
- Ecologically valid sample recruited in statistics classes by their statistics instructors.
- Visuospatial working memory best predicts mental manipulation in spatial anxiety, which best predicts statistics interpretation anxiety.
- Reached large samples, but probably due to the COVID-19 pandemic and technological limitations, the response rates were lower.
- spatial skills best predict mental manipulation subscale of spatial anxiety, which best predicts interpretation of statistics anxiety.

Disclaimer

Minor changes were made for the present thesis that are not in the published article to improve the text based on the comments of the thesis committee members.

Abstract

The present two studies investigated the role of spatial cognition on statistics anxiety. The hypothesis that spatial representations and/or visuospatial skills are related to the acquisition of statistics abilities which, when lacking or unused, generate statistics anxiety is examined. To this end, a total of 680 students in social sciences from 14 different universities located in one of three countries enrolled in a statistics class at the time of the study were recruited. Participants were mostly women. Study 1 examined a mediation model where visuospatial and verbal working memory spans as well as spatial anxiety are predictors of statistics anxiety with mathematics anxiety as the mediator. The results show a partial mediation and strong associations between all three types of anxiety (i.e., spatial anxiety, mathematics anxiety and statistics anxiety). The subscale *statistics interpretation* anxiety was best predicted by visuospatial working memory span. Study 2 examined a path regression model where performance on a spatial and a verbal task along with spatial anxiety are predictors of statistics anxiety. The results indicate that the *mental manipulation* subscale of spatial skills is a strong predictor of mental manipulation anxiety which, in turn, predicts interpretation anxiety in statistics. Both studies support the role of spatial cognition in statistics understanding. These results have implications for the teaching and learning of statistics.

Keywords: Statistics anxiety, spatial anxiety, mathematics anxiety, working memory, spatial skills

There is widespread agreement that statistics education needs to be improved (e.g., Carver et al., 2016; Moreira Da Silva & Pinto, 2014). Difficulties in learning and teaching statistics stem from a multitude of factors (Cousineau & Harding, 2017; Cui et al., 2019; Harth, 2018). On the learners' side, a major factor impeding the learning of statistics is students' fears, worries and apprehensions toward any form of statistical concepts or operations (Gibeau et al., 2023; Onwuegbuzie et al., 1997). This well-documented phenomenon is known as *statistics anxiety* (Esnard et al., 2021; Ruggeri et al., 2011; Williams, 2015; Zeidner, 1991).

This construct is multi-faceted. However, there is no consensus on the number of facets and their nature (see Study A of Chapter 2 of the present thesis for more detail; Cruise et al., 1985, Onwuegbuzie et al., 1997, Vigil-Colet, 2008; Zeidner, 1991). Herein, we use the three-dimension model (Cantinotti et al., 2017; Gibeau et al., 2023; Vigil-Colet et al., 2008). The first dimension concerns anxiety toward exams and tests in a statistics course (*evaluation anxiety*). The second dimension is about asking the instructor or the teaching assistants for help in understanding something or doing an exercise (*anxiety of asking for help*). Lastly, *interpretation anxiety* concerns anxiety toward interpreting statistical results. This framework comes from the work of Vigil-Colet et al. (2008), in which they developed the *Statistical Anxiety Scale*. It is the framework used herein as it has good factorial structure compared to the factor structure of a psychometric evaluation study of the other leading framework in statistics anxiety; the *Statistical Anxiety Rating Scale* (STARS; Nesbit & Bourne, 2018; originally developed by Cruise et al., 1985).

Statistics anxiety is widely observed in university students (Chiou et al., 2014). In particular, Zeidner (1991) reported a high level of statistics anxiety in 70% of social sciences students. Similarly, Onwuegbuzie and Wilson (2003) reported that a majority of graduate students felt "uncomfortable" with statistics. A more recent study reported that half of a large sample of college students enrolled in a wide range of programs have moderate to severe statistics anxiety (Huang et al., 2023). According to previous research, statistics anxiety is moderately related to grades in statistics courses. Specifically, Cantinotti et al., (2017) reported a correlation of -.33 between these two variables (see also Fullerton & Umphrey, 2016, Chiesi & Primi, 2010, and Mackinnon et al., 2025, for similar findings); making this variable one of the strongest predictors of the final grade and presumably, an important predictor of statistics *learning*. Identifying the determinants of statistics anxiety and better understanding how it relates to achievement in statistics could facilitate the development of learning activities to improve students' understanding of the underlying concepts of statistics. Indeed, as statistically literate citizens, people will be better qualified to contribute to societal discussions and debates in a world where policies need to be informed by relevant data (e.g., Cui et al., 2019; Lewin et al., 2012).

A few reliable relations between statistics anxiety and other constructs are now well established, such as gender (Edirisooriya & Lipscomb, 2021; Mandap, 2016, but see Gibeau et al., 2023, which question this relation), achievement goals (Lalande et al., 2019; Valle et al., 2021), attitudes toward the subject (Chiesi & Primi, 2010; Macher et al., 2013; McIntee et al., 2022) and mathematics anxiety (Baloğlu, 2004; Jordan et al., 2014; Zeidner, 1991). In the present studies, three cognitive factors that might be *predictors* of statistics anxiety are examined. Those are working memory, spatial anxiety and mathematics anxiety.

Potential Predictors of Statistics Anxiety

Research involving statistics anxiety and statistics ability is only now emerging. Indeed, there is not much research on the influence of working memory on statistics. As such, this subsection considers research on mathematics anxiety or ability as they are strongly related but distinct constructs (Baloğlu, 1999, 2004; Zeidner, 1991).

The Influence of Working Memory in Mathematics

Working memory is strongly related to mathematics (Raghubar et al., 2010; van de Weijer-Bergsma et al., 2022). Indeed, the understanding of a problem cannot be better than its internal representation in working memory. Specifically, Raghubar et al., (2010) argue that working memory, by definition, is related to mathematical skills since “problem solving in these domains often involves the holding of partial information and the processing of new information to arrive at a solution” (p.110). This association is evidenced in multiple studies reviewed in their article. For example, a previous study indicates that the Phonological Loop and the Visuo-Spatial Sketchpad (two of three working memory components according to Baddeley, 1992) influences performance on a multiplication task and a subtraction task, respectively, when measured simultaneously (Lee & Kang, 2002).

Similarly, Giofrè et al. (2018) observed a distinction between verbal working memory span and spatial working memory span with regards to mathematics and statistics. Specifically, the authors tested three models: one in which both memory spans were considered undistinguishable onto the memory tasks; a second model in which both memory spans were considered separately onto the memory tasks; and a third in which both memory spans were considered separately along with mathematics and reading ability onto their corresponding tasks. The second model had a better fit than the first one, suggesting that verbal working memory and visuospatial working memory are involved in distinct cognitive processes. The authors reported that visuospatial working memory explained 10% of the variance in mathematics while verbal working memory explained only 4%.

This supports the hypothesis that visuospatial working memory, compared to verbal working memory, should be more related to the present study's main variables (spatial, mathematics and statistics anxieties). Furthermore, the literature on the relation between working memory and anxiety is extensive, specifically with mathematics anxiety. For example, a recent study looked at verbal working memory, numerical working memory, mathematics anxiety and calculation performance (Peng & Ma, 2025). The results suggest that only verbal working memory is a significant mediator in the relation between mathematics anxiety and calculation performance. However, the authors only examined the phonological loop component of working memory using two different stimuli. It is still unclear whether visuo-spatial working memory (the other component of the working memory system according to Baddeley, 2010). In another recent study, the authors included spatial working memory (He et al., 2025). The results suggest that children with high spatial and verbal working memory, tend to be less mathematical anxious, and those with mathematics anxiety and low spatial working memory have a lower performance in an approximate number system task. It is consequently hypothesized that statistics anxiety will be more negatively associated with visuospatial working memory span than with verbal working memory span (which will serve as a control).

Only one study, that we know of, examined working memory and mathematics alongside statistics, instead of only mathematics. Indeed, Bonnot and Croizet (2007), looking at the relations between statistics performance, mathematics performance and working memory related to gender differences, concluded that women who held the inferiority stereotype in mathematics also had lower grades in statistics and lower self-evaluation in mathematics. Additionally, there was a bigger difference in mathematics performance when working memory was involved. In other words, women who held the inferiority stereotype had lower mathematics performance and lower self-evaluation only in difficult tasks (i.e., when working memory is more solicited). These studies suggest that working memory may be associated to statistics ability, mathematics ability and spatial skills (described below).

The Spatial Skills Influence on Statistics Anxiety

Spatial skills are cognitive abilities that allow one to mentally manipulate, transform and represent elements within their mind (Bennett-Pierre & Gunderson, 2023; Uttal et al., 2013). They are typically grouped into one of three dimensions: *mental manipulation*, *navigation*, and *imagery* (Lyons et al., 2018). The first dimension involves crude and dynamic elements, the second involves dynamic and vague or detailed elements, and the last one involves static elements with many details.

These skills are believed to be used to solve and accomplish a wide variety of problems and tasks (Davis et al., 2021; Jaeger et al., 2016; Miller et al., 2020). For example, spatial skills appear critical to mathematics learning, understanding, and performance (Geary et al., 2023; Kyttälä & Björn, 2014; Rittle-Johnson et al., 2018). Specifically, a recent study suggests that spatial skills and attention during class are the two best predictors of mathematics performance (Geary et al., 2021). Furthermore, spatial skills are related to problem-solving and reasoning activities required in many professions, notably those that are STEM-related (Science, Technologie, Engineering and Mathematics; Atit et al., 2022; Sorby et al., 2018). This relation between spatial skills and competence and interest in STEM fields has been amply described (Hinze et al., 2013; Sorby et al., 2018; Yang et al., 2020). However, studies examining the relations between spatial skills and statistics learning in non-STEM fields are only now emerging.

In recent work, statistics anxiety has been linked to the anxiety toward activities or reasonings requiring spatial skills (*spatial anxiety*; e.g., anxiety toward needing to mentally manipulate an element; Gibeau et al., 2023). Specifically, it was observed that people higher in mental manipulation anxiety are also higher in statistics interpretation anxiety, or vice versa. Indeed, interpretation of statistics anxiety was most strongly positively associated to mental manipulation anxiety, weakly positively associated with imagery anxiety and not at all with navigation anxiety. Cousineau and Harding (2017) argued that the mental representations underlying statistical reasoning are visuospatial in nature. This hypothesis explains why statistics anxiety relates to spatial anxiety (Gibeau et al., 2023). Herein, this hypothesis is explored more fully by examining actual spatial *skills* in addition to spatial *anxiety*.

Even though there is an established relation between spatial anxiety and statistics anxiety, it does not necessarily follow that spatial performance will also relate to statistics anxiety. In the current studies, the goal is to deepen the understanding of the link between statistics anxiety and spatial processing by examining the relations between performance on measures of spatial skills and responses to measures of spatial anxiety. Specifically, the relations between performance on two separate measures of spatial skills are explored: visuospatial working memory span (Rouet et al., 2012; Wang et al., 2021) and spatial problem solving (Henn et al., 2018; Rodán et al., 2022). Study 1 examines spatial processing on statistics anxiety using visuo-spatial working memory. Study 2 examines spatial processing on statistics anxiety using spatial problem solving.

Hypotheses

The main goal of this study is the examine the role of spatial processing skills (through visuo-spatial working memory and then through spatial problem solving) in statistics anxiety, while considering spatial anxiety and mathematics anxiety. First, it is hypothesized that visuo-spatial working memory, compared to verbal working memory, influences the levels of spatial anxiety, which, in turn, will influence statistics anxiety through mathematics anxiety. It is also hypothesized that spatial skills, compared to verbal abilities, influence the levels of spatial anxiety, which, in turn, will influence statistics anxiety through mathematics anxiety. Below is the explanation of how theses hypotheses came to be.

Mathematics anxiety will be used as a mediator between spatial anxiety and statistics anxiety. This decision is motivated by the possible presence of a cascade of influence between those three variables. Specifically, spatial anxiety appears quite early in children (between 6 to 8 years old; Lauer et al., 2018; Ramirez et al., 2012) and mathematics anxiety has been reported around 8 to 9 years of age (fourth grade; Harari et al., 2013). In their study, Harari et al. (2013) reported moderate mathematics anxiety in first graders (6-7 years old); however, it correlated weakly with performance in mathematics. Statistics anxiety is thought to appear later as it is taught quite late in students' curriculum as a separate subject (i.e., college or university). Developmental longitudinal studies are needed to formally test this cascade of influence. Nevertheless, because spatial skills predict mathematics anxiety and mathematics ability, as described above, they may also influence statistics anxiety and statistics ability. To summarize, it is hypothesized that spatial skills negatively relate to spatial anxiety, which directly and indirectly (through mathematics anxiety) relates to statistics anxiety. This is explored in two studies.

In Study 1, visuospatial working memory span is measured with verbal working memory span as a control. The relation of visuospatial working memory span and statistics anxiety should be different for all three subscales of statistics anxiety, and is expected to be stronger for statistics interpretation anxiety relative to the other two, as spatial anxiety was more positively associated with statistics interpretation anxiety in a previous study (Gibeau et al., 2023). In

Study 2, spatial skills are measured through a spatial reasoning test. It is hypothesized that the same pattern of associations will occur when spatial skills are used and when verbal ability is used as a control in a path regression model, similarly to numerous previous studies involving mathematics anxiety (e.g., Delage et al., 2021; Ferguson et al., 2015). When looking at the items measuring the anxiety of asking for help in the *Statistical Anxiety Scale* (SAS), one can see that the statements have a social component as it involves another person (e.g., “Asking the teacher how to use a probability table”). This is why the anxiety of asking for help alongside general anxiety are considered as controls in the present research.

Study 1

Models of working memory (WM) propose two specialized short-term memory systems, one for manipulating phonological information (the *phonological loop*) and a second for manipulating visuospatial information (the *visuospatial sketchpad*; Baddeley, 1992). One measure for each of those specialized dimensions is used in the present study, the Corsi-Blocks task, and the n-Back task. Baddeley (1992) also assumed a central executive common to both memory systems.

Method

Participants

Participants were students recruited from Canadian, French, and Belgian universities. All participants were asked to complete the survey in French. They were undergraduates in social sciences or human sciences programs and were enrolled in an introductory statistics course at the time of the study. They were informed of the study by their statistics instructor. Data collection was planned to occur during the academic year 2020–2021 but it was extended for an additional academic year (2021–2022; see Results section). Data were collected during the fall and winter semesters during these two academic years. No exclusion criteria were applied during or prior to data collection. The instructors were invited to use a one-page slide promoting the study, a short video showing the first and last author explaining the study, and/or an email sent to their students (according to their preference). It was specified to students that their course grade would not be affected by their decision to participate or not and that their instructor would not be given information about their participation.

The target sample size was 800 based on a previous study on the same population (Gibeau et al., 2023). However, only 306 participants were included in the final analyses. Among those 306 participants, 255 identified as women, 43 as men, and 8 answered “other,” “prefer not to say” or did not answer the question. A total of 240 participants answered the question about their age. The majority of those 240 participants are from 18 to 24 years-old, with a mean of 22 and a median of 20. Participants were recruited from two different continents: 213 from Europe (either France or Belgium) and 93 from North America (Canada).

Materials

Questionnaires

Spatial Anxiety. Spatial anxiety was measured with the validated French version of the *Spatial Anxiety Questionnaire* (SAQ-F; Gibeau et al., 2023; original version in Lyons et al., 2018). All the items are preceded by a context and a question, “Are you anxious when…”. This instrument has three subscales: mental manipulation anxiety (SAQ-M; e.g., “… you are asked to determine how a series of pulleys will interact given only a 2-dimensional diagram”), navigation anxiety (SAQ-N; e.g., “… you are asked to do the navigational planning for a long car trip”) and imagery anxiety (SAQ-I; e.g., “… you are asked to recall the shade and the pattern of a person’s tie you met for the first time the previous evening”). Each subscale included eight items, for a total of 24 items. Participants indicated how much anxiety they would experience if facing the situation described on a five-point scale (0: “not at all,” 1: “a little,” 2: “a fair amount,” 3: “much,” 4: “very much”). The range of possible scores is 0-96, with higher scores indicating higher anxiety. The authors of the French validation study reported excellent internal reliability (ω = .93 overall). Convergent validity of the original version was assessed by the relations between its three subscales and their corresponding attitudes and skills while controlling for general anxiety (Lyons et al., 2018). The authors reported that each subscale was moderately-to-strongly related with its relevant self-reported and objective skill.

Statistics Anxiety. This variable was measured with the validated 21-item French version of the *Statistical Anxiety Scale* (SAS-F; Cantinotti et al., 2017; original version in Vigil-Colet et al., 2008). The original version included 24 items divided into three subscales: evaluation anxiety (SAS-E; e.g., “… studying for a statistics exam”), anxiety of asking for help (SAS-A; e.g., “… asking the professor or their assistant(s) to explain a concept I did not understand”) and interpretation anxiety (SAS-I; e.g., “… reading advertisement that includes tables and graphs”). As reported in Cantinotti et al. (2017), three items did not contribute enough to the confirmatory factor analysis on the French version and were removed. Again, participants indicated how much anxiety they would experience if facing the situation described in each item on a five-point scale; “no anxiety” (1), “…” (2), “medium anxiety” (3), “…” (4) and “a lot of anxiety” (5). The range of possible scores is 21-105, with higher scores indicating higher anxiety. The authors of the French version reported excellent reliability of scores (ω = .95 overall). Validity was assessed by a clear 3-factor structure and by a strong correlation with students’ final grade in a statistics course (Cantinotti et al., 2017). However, Mackinnon et al. (2025) concluded that grades in a statistics course predict higher statistics anxiety.

Mathematics Anxiety. This variable was measured with the validated French version of the *Abbreviated Math Anxiety Scale* (AMAS; Gibeau et al., 2023; original version in Hopko et al., 2003). The questionnaire measures mathematics anxiety on two dimensions: learning mathematics anxiety (AMAS-L; e.g., "… listening to a lecture in math") and mathematics evaluation anxiety (AMAS-E; e.g., "… you receive a pop quiz in your math class"). The learning anxiety subscale has five items, and the evaluation anxiety subscale has four items, for a total of nine items. Participants indicated how much anxiety they would experience if facing the situation described on a five-point scale; "low anxiety" (1), "…" (2), "medium anxiety" (3), "…" (4) and "high anxiety" (5). The range of possible scores is 9-45, with higher scores indicating higher anxiety. The authors of the French validation study reported excellent reliability of scores ($\alpha = .90$ overall). Validity is indicated by a strong correlation with the statistics anxiety scale ($r = .79$; Gibeau et al., 2023).

Tasks

Tasks were programmed in JavaScript and implemented into Qualtrics.

n-Back Task. This task measures executive functions as well as phonological working memory span (Baddeley, 1992). A previous study examined the task's validity (Kearney-Ramos et al., 2014). The authors examined the task's convergent and divergent validity using correlations with other neuropsychological assessments (i.e., Digit Span Test, Spatial Span Test, Test of Everyday Attention and Delis-Kaplan Executive Function System Trail-Making Test for convergent validity; Halstead-Reitan Finger Tapping Test and Boston Naming Test-2 for divergent validity). The authors concluded that this is a valid measure for working memory because of both significant and non-significant correlations. Reliability was measure by Kulikowski and Potasz-Kulikowska (2016), using split-half Pearson's correlations (*r*s between

.61 and .83 with participants' accuracy). The results indicate acceptable-to-good reliability. One trial featured a series of letters with alternating upper- and lower-case to avoid similarity-based recall (e.g., B, f, H, k, M, q, R, x; Vytal et al., 2013). They were presented in the center of the screen, one letter at a time, with a font size of 30. They were all black over a white background. The first three letters never included targets. All letters were presented for 1,000 ms with an inter-item delay of 500 ms. Each trial was composed of 30 items including 10 targets.

Participants were asked to press on the space bar when the letter displayed on the screen was the same as the letter presented just before (1-back condition), two letters before (2-back condition) or three letters before (3-back condition). The 1-back condition is trivially easy for the intended population. Each of the three conditions was repeated three times. Conditions were presented in ascending order of difficulty. The number of correct responses in each condition (a correct response being either a correct spacebar press or a correct no-response) were recorded for a score from 0 to 90 per n-back condition. A participant who never presses the space bar obtains a score of 60 out of 90 (i.e., 30 errors). The score was rescaled on a range from 0 to 100. If the cumulative number of errors reached 24 or more at the end of a condition, the task was interrupted and the participant could not continue to the next condition (see the implementation given in Bradley-Garcia & Bolton, 2023).

Corsi-Blocks Task. To measure participants' ability to memorize visuospatial information, the forward Corsi-Blocks Task was used (Corsi,1972; Milner, 1971). Validity and reliability evidence for measuring spatial working memory is difficult to find as there are many different versions of the task (Gibeau, 2021). Note that there are many measures of visuospatial working memory. It was considered convenient for the present study in terms of accessibility and programmability. Negative moderate correlations with all three anxiety measures used herein (see Table 5) are evidence of good convergent validity. Internal consistency reliability was tested by de Paula et al. (2016), using a split-half method. The results indicate acceptable reliability (0.75). Note that the specific statistic used to estimate the internal consistency is not mentioned.

Nine black squares on a white background were presented on participants' screen at random locations. The size of each square was 10% of the screen's width and the minimal distance between the squares was 1% of the screen's width. A sequence of highlighted squares was presented. To highlight the squares, they turned from black to green for 400 ms. The sequences are of varying lengths; they can be composed of three to seven squares with four trials of each for a total of 20 trials with an inter-trial delay of 2,000 ms. During the presentation of the sequence, participants' mouse cursor was hidden to avoid motor recall. At the end of the sequence, to indicate that the recall period has begun, a 500 Hz audio signal was played for 400 ms with the volume set at 15% of the highest volume. To recall the sequence, participants needed to click on the correct squares in the correct order. When participants clicked on a square, it turned gray for 200 ms (see the implementation given in Gibeau, 2021). A trial was coded as successful (1 point) if the entire sequence was reported in the correct order. Sequences of length 3 are very easy to perform for healthy adults. Each condition of that task —sequences from 3 to 7 squares— received a score between 0 and 4 (as there were four trials per condition). Then, the score of each condition was weighted so that success in recalling 7 squares (multiplied by 5) was worth more than success in recalling 3 squares (multiplied by 1). The resulting score, ranging from 0 to 60, was rescaled on a range from 0 to 100 to ease interpretation.

Procedure

In the promotional material, a URL link to an online Qualtrics study was distributed. The study took approximately 30 minutes to complete. No compensation was provided, but Canadian participants could enter into a draw to win a $50 gift card. One gift card for every 50 participants was drawn at the end of each academic year.

First, participants provided informed consent and answered a short series of socio-demographic questions. Then, they completed three questionnaires followed by two tasks as described above. The questionnaires and their items were presented in a randomized order.

The items were divided into seven blocks. The first four blocks contained eight items from the *Spatial Anxiety Questionnaire* (SAQ) and eight from the *Statistical Anxiety Scale* (SAS; two blocks for each questionnaire in a random order). The fifth and sixth blocks contained the last eight items of the SAQ and the last five items of the SAS. Finally, the last block was composed of the nine items of the *Abbreviated Math Anxiety Scale* (AMAS). Each questionnaire's total score as well as its subscale scores were rescaled on a range from 0 to 100 to ease the interpretation. As recommended by Cousineau (2020), descriptive statistics are reported with one decimal and standardized effect sizes are reported with two decimals in the present study.

Results

Data Screening

In total, 1,332 people clicked on the URL link and 420 completed the questionnaire (31.5%). This was unexpected as in a prior study (September 2019-April 2020) involving the same population and the same universities, 82.1% of the individuals who clicked on the URL completed the whole survey (Gibeau et al., 2023). One possible explanation may be that the COVID-19 pandemic influenced the students, making them too preoccupied and more prone to leave the study. Another explanation might be that the tasks were judged as too demanding. Furthermore, many of the remaining participants (114) had a score of 0 in all three conditions of the Corsi-Blocks Task (presumably due to browser or device incompatibility with the program) and were excluded from further analyses. Thus, a total of 306 participants were included in the final analyses. This is the reason why data collection was extended over a second academic year. Among the remaining data, there were 105 missing values (0.6% of the items). The Little test was performed to make sure that the missing data were missing completely at random ($\chi^2 = 1689.55$, $df = 1621$, $p = .12$).

All participants completed the 1-back task (90 without error), but 41 (13%) did not progress to the 2-back condition. Of the remaining 265 participants, almost all (262) reached the 3-back condition. Skews and kurtosis were high in the 1-back condition (-2.02 and 3.45 respectively) and in the 2-back condition (-1.67 and 2.85 respectively). This might be due to a ceiling effect, where the first two conditions are too easy. For the last condition (3-back), skew and kurtosis were acceptable (-0.41 and -0.31 respectively). An arcsine transformation on the proportion correct for each condition was used (Laurencelle & Cousineau, 2022), which considerably reduced these deviations from normality (the skews are -1.39, -0.79, and 0.12, respectively, and the kurtosis are 0.99, 0.13, and -0.40, respectively). The total score to the n-Back Task was the sum of the arcsine-transformed scores on the three conditions, rescaled on a range of 0 to 100.

The Corsi-Blocks Task was easy for sequences of 3 and 4 squares (mean number of successful trials of 3.8 and 3.5 out of 4 trials) then accuracy decreased abruptly for sequences of 5 to 7 (means of 2.8, 1.8 and 0.9). Asymmetry was initially negative (too easy with a ceiling effect) then became positive (too hard with a floor effect), ranging from -3.40 to 0.95. Kurtosis also varied significantly across conditions (ranging from 11.95 to -1.02). However, the weighted total score, used for further analyses, was nearly normally distributed (skew of -0.09 and kurtosis of -0.35).

Descriptive Statistics

Table 1 shows the descriptive statistics of participants' tasks, and questionnaires. A prior study found no between-continent differences with respect to the questionnaires used herein (Gibeau et al., 2023), thus the present analyses are collapsed across continents. Finally, the questionnaires' reliability indices were very similar to those reported in the original studies (McDonald's omega ranging from .92 to .96).

Table 1

Descriptive Statistics of all Variables used in Study 1

Variables	*N*	Mean [95% CI]	SD	Min.	Max.	Skewness	Cronbach's α	McDonald's ω
NBack	306	77.6 [76.0, 79.2]	14.1	38.8	94.7	-1.54	-	-
CBlocks	306	51.8 [49.6, 53.9]	19.5	2	100	-0.09	-	-
SAQ	306	42.7 [40.4, 44.5]	18.0	0	91	0.80	.92	.94
SAS	306	50.8 [48.3, 53.1]	20.1	0	96	-0.09	.93	.96
AMAS	306	46.8 [44.1, 49.4]	22.6	0	100	0.24	.90	.92

Notes. NBack signifies the *n-Back Task*; CBlocks signifies the *Corsi-Blocks Task*; SAQ signifies the *Spatial Anxiety Questionnaire*; SAS signifies the *Statistics Anxiety Scale*; AMAS signifies the *Abbreviated Math Anxiety Scale*. All the questionnaires are using the French-validated versions. The questionnaires and tasks were rescaled on a range from 0 to 100 points.

Correlations Between Questionnaires' Total Scores and Tasks

First, the relations between the three anxiety questionnaires were examined (see Table 2). Replicating the results of Gibeau et al. (2023), all three anxieties were moderately-to-strongly positively correlated (r = .49 to .74). As expected, mathematics anxiety and statistics anxiety show the largest correlation (r = .74). These correlations, as well as the following correlations in the present text, were compared using the `Medcalc.org` website. The comparisons of these correlations are significant. The result of the comparison between the spatial anxiety and mathematics anxiety correlation (r = .49), and the mathematics anxiety and statistics anxiety correlation (r = .74) is: $z = -5.10$, $p < .0001$. As for the comparison between the spatial anxiety and statistics anxiety correlation (r = .52), and the mathematics anxiety and statistics anxiety (r = .74), the results are: $z = -4.61$, $p < .0001$. The two working memory span tasks were weakly positively associated (r = .15), indicating that they do not measure the same components of working memory.

Next, the hypothesis that performance on the Corsi-Blocks Task would be negatively related to each of the anxiety questionnaires was tested. As predicted, there was significant negative correlations between performance on the Corsi-Blocks Task and the anxiety questionnaires (spatial anxiety r = -.13, mathematics anxiety r = -.24 and statistics anxiety r = -.13). Performance on the n-Back Task was also negatively correlated with the questionnaires (spatial anxiety r = -.10, mathematics anxiety r = -.14 and statistics anxiety r = -.13). Even though the correlations are weak to moderate, they have the expected directionality. By contrast, the correlations between scales' totals are strong because they are all quite similarly framed, asking questions about self-reported anxiety.

Table 2

Correlation Matrix of all Questionnaires' total scores and Tasks in Study 1

Variables	NBack	CBlocks	SAQ	SAS
CBlocks	.15 [.05, .25]			
SAQ	-.10 [-.21, .02]	-.13 [-.21, -.02]		
SAS	-.13 [-.25, -.02]	-.13 [-.25, -.02]	.52 [.42, .61]	
AMAS	-.14 [-.26, -.03]	-.24 [-.35, -.11]	.49 [.39, .59]	.74 [.69, .79]

Notes. NBack signifies the *n-Back Task*; CBlocks signifies the *Corsi-Blocks Task*; SAQ signifies the *Spatial Anxiety Questionnaire*; SAS signifies the *Statistics Anxiety Scale*; AMAS signifies the *Abbreviated Math Anxiety Scale*. The numbers between brackets are the 95% confidence intervals of the correlation coefficient.

Correlations Between Anxiety Subscales

Relations between the components of the three anxieties were explored. Mathematics anxiety subscales (AMAS-L and AMAS-E; last two rows of Table 3) were strongly correlated with statistics anxiety subscales (SAS-E, SAS-A, and SAS-I). These high intercorrelations were expected and are frequently observed for these two related forms of anxiety (Paechter et al., 2017). Anxiety of asking for help in statistics had the weakest correlations with both subscales of mathematics anxiety (.37 and .38 respectively; comparison of AMAS-L – SAS-E and AMAS-L – SAS-A: $z = -8.95$, $p < .0001$; comparison of AMAS-L – SAS-I and AMAS-L – SAS-A: $z = -2.51$, $p = .01$; comparison of AMAS-E – SAS-E and AMAS-E – SAS-A: $z = -2.48$ $p = .01$; comparison of AMAS-E – SAS-I and AMAS-E – SAS-A: $z = -6.14$, $p < .0001$). This was unsurprising as the AMAS has no explicit social dimension. The fact that mathematics evaluation anxiety was less related to statistics evaluation anxiety than to statistics interpretation anxiety is an oddity observed in other studies (e.g., Gibeau et al., 2023).

Table 3

Correlation Matrix of Subscales and Tasks Scores in Study 1

Variables	NBack	CBlock	SAQ-M	SAQ-I	SAQ-N	SAS-E	SAS-A	SAS-I	AMAS-E
CBlock	.15								
	[.04, .26]								
SAQ-M	-.13	-.19							
	[-.24, -.02]	[-.30, -.07]							
SAQ-I	-.02	.04	.44						
	[-.14, .10]	[-.07, .16]	[.34, .54]						
SAQ-N	-.07	-.14	.44	.44					
	[-.18, .04]	[-.25, -.02]	[.35, .54]	[.33, .54]					
SAS-E	-.10	-.20	.42	.23	.35				
	[-.23, .04]	[-.33, -.06]	[.32, .52]	[.11, .34]	[.23, .45]				
SAS-A	-.10	≈ .00	.26	.29	.32	.40			
	[-.21, .01]	[-.11, .10]	[.16, .36]	[.17, .39]	[.21, .42]	[.30, .49]			
SAS-I	-.12	-.16	.55	.29	.28	.56	.40		
	[-.23, -.01]	[-.27, -.04]	[.47, .63]	[.18, .39]	[.18, .39]	[.50, .62]	[.30, .49]		
AMAS-E	-.14	-.24	.49	.24	.21	.53	.37	.71	
	[-.25, -.02]	[-.33, -.14]	[.38, .58]	[.11, .36]	[.10, .32]	[.46, .58]	[.28, .47]	[.63, .77]	
AMAS-L	-.12	-.18	.45	.31	.38	.81	.38	.54	.62
	[-.24, .01]	[-.30, -.06]	[.35, .55]	[.19, .42]	[.28, .48]	[.76, .85]	[.27, .48]	[.46, .62]	[.56, .67]

Notes. NBack signifies the *n-Back Task*; CBlocks signifies the *Corsi-Blocks Task*; SAQ-M signifies the Mental manipulation subscale of the *Spatial Anxiety Questionnaire*; SAQ-I signifies the Imagery subscale of the *Spatial Anxiety Questionnaire*; SAQ-N signifies the Navigation subscale of the *Spatial Anxiety* Questionnaire; SAS-E signifies the Evaluation Anxiety of the *Statistics Anxiety Scale*; SAS-A signifies the Fear of Asking for Help subscale of the *Statistics Anxiety Scale*; SAS-I signifies the Interpretation Anxiety subscale of the *Statistics Anxiety Scale*; AMAS-L signifies the Learning Anxiety subscale of the *Abbreviated Math Anxiety Scale*; AMAS-E signifies the Evaluation Anxiety of the *Abbreviated Math Anxiety Scale*. The numbers between brackets are the 95% confidence intervals of the correlation coefficient.

Regarding associations between the three spatial anxiety subscales and the five other questionnaires' subscales, all correlations were strong, and their distribution seemed bimodal. Eleven of them were around .30 (from .20 to .38 with a mean of .29) whereas four were above .42 (.42 to .55 with a mean of .48). All four of them were with respect to SAQ-M. This is interpreted as indicating that a correlation of about .30 is a baseline for associations between anxiety variables. Hence, any correlation above .30 denotes a substantive association based on the cognitive processes. Note that this conclusion is based solely on the correlations reported here and needs further research. The lack of a substantive correlation between SAQ-M and anxiety of asking for help ($r = .26$) supports this interpretation. The numerically, but not statistically, strongest correlation, $r = .55$, relates SAQ-M to SAS-I, in line with the main hypothesis.

Correlations Between Anxiety Subscales and Working Memory Tasks

Performance on the n-Back Task was weakly and negatively associated with all eight subscales of the anxiety scales (*rs* from -.14 to -.02). The correlations were all negative, suggesting that participants with good phonological span or good executive functions experience lower anxiety, and vice versa. The average r was -.10, with 4 of the 8 correlations involving anxiety subscales which did not differ significantly from zero. Notably, the correlation between the n-Back Task and anxiety of asking for help was -.10. This last variable was a control as it is considered more social in nature, meaning that it is considered as an indication of social anxiety.

The Corsi-Blocks Task showed more varied correlations (ranging from -.24 to ≈ .00). It was null with respect to anxiety of asking for help and very weak for imagery anxiety. The Corsi-Blocks Task involves schematized blocks (mere squares), which does not require any highly detailed imagery. As such, imagery anxiety was not involved in this task, or vice versa.

The remaining six correlations (ranging from -.24 to -.14 with a mean of -.19) were all significant (all *p*s < .015). This critical result suggests that visuospatial working memory span has some relation with anxiety toward spatial tasks, mathematics anxiety, and anxiety of statistics evaluation and statistics interpretation (but not toward anxiety of asking for help).

Although the correlations involving the Corsi-Blocks Task were not very strong, they were almost twice as large as the correlations involving the n-Back Task. Thus, even though the n-Back Task has an advantage (it measures two constructs: phonological span and executive functions), it failed to capture a significant part of the variance, and in addition, it did not distinguish social anxiety from interpretation anxiety. Consequently, it lacks specificity when considering statistics and mathematics.

The two critical correlations regarding the main hypothesis were the Corsi-Blocks Task and the n-Back Task on interpretation anxiety in statistics. The correlation is strongest for the Corsi-Blocks Task (r = -.16 vs. r = -.12) but the difference was small and not significant (z = .50, p = .62). One reason is that these variables may have variance which overlaps with other sources of variance.

To determine whether they are distinct correlations from a unique source of variance, semi-partial correlations were computed to find the specific portion of variance explained by the tasks over and above anxiety of asking for help (Abdi, 2007; Cohen & Cohen, 1975). Notably, each task had non-significant correlations with anxiety of asking for help. As this latter measure is a control, it was used to discount the contribution of anxiety items in general from the interpretation anxiety in statistics. Table 4 repeats the (regular) correlations and shows the semi-partial correlations for all three predictors with respect to interpretation anxiety in statistics. These correlations were run post-hoc to see the unique contribution of the Corsi-Blocks task when the other predictors are considered.

Table 4

Pairwise and semi-partial correlations for three predictors of interpretation anxiety in statistics (SAS-I) in Study 1

	Correlations	
Variables	Pairwise	Semi-Partial
NBack	-.12	-.06
CBlock	-.16	-.16
SAS-A	.40	.40

Note. The pairwise correlations are reproduced from Table 3 for easier comparison.

As seen, the correlation between the n-Back Task and interpretation anxiety in statistics, when considering anxiety of asking for help and the Corsi-Blocks Task, decreased almost two-fold whereas the other two stayed nearly unchanged. This indicates that the n-Back Task shares variances with the other two variables whereas the Corsi-Blocks Task shares none with anxiety of asking for help (as seen in Table 3, this last correlation was undistinguishable from zero).

In the next section, two models are examined. These further suggest a stronger, distinct, contribution of visuospatial working memory span onto interpretation anxiety in statistics, compared to phonological working memory span.

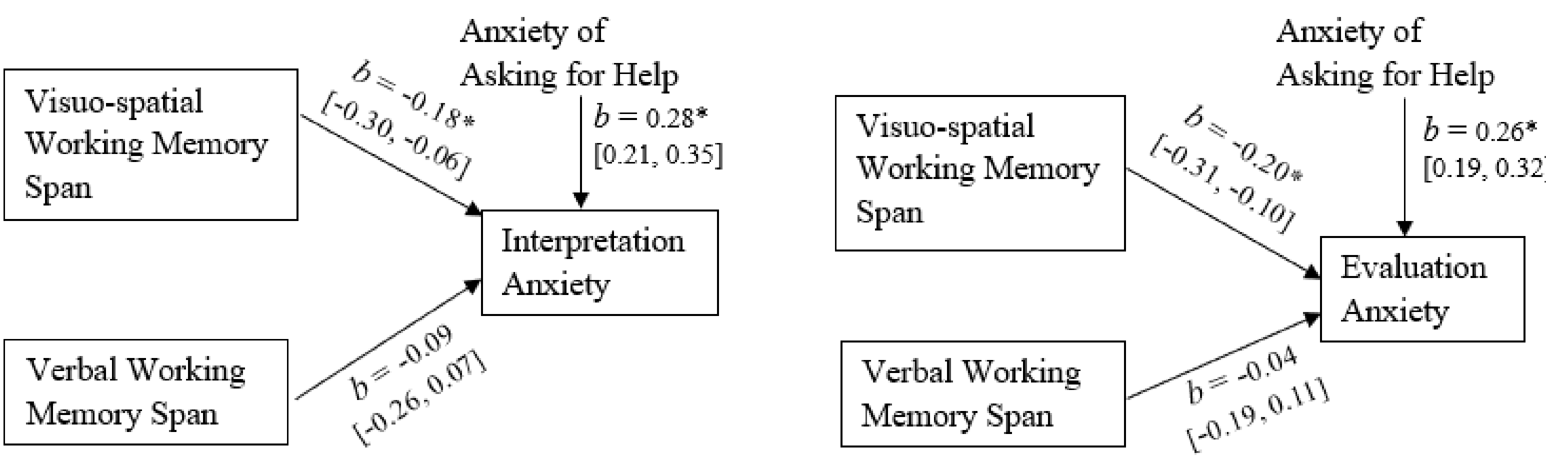
Path Regression Models

A post hoc power analysis was run in GPower to determine the statistical power for this analysis. With an effect size of 0.18 (the result of the first path regression from Figure 1), an alpha of .05, a sample size of 306 and three predictors, the power was estimated at .99.

When examining the correlations in the sections above, it was argued that there is a baseline association involving measures of anxiety. Hence, to get the distinct contribution of working memory spans onto interpretation anxiety in statistics, it is necessary to control for that baseline. This is what covariance analysis does. Furthermore, considering that gender could influence the results and that the present sample is mostly women, it is also included as a covariate. Here, a linear regression predicting interpretation anxiety in statistics while controlling for anxiety of asking for help was performed using the mediation command in the MPlus software (version 7). Figure 1, left-hand panel, shows the results. As can be seen, the role of visuospatial working memory span was stronger when the baseline is controlled for. As an exploratory analysis, the same regression was run with evaluation anxiety in statistics. The results show a similar pattern (Figure 1, right-hand panel).

Figure 1

Regression of Corsi Blocks and n-Back tasks when anxiety of asking for help is used as a covariate on interpretation anxiety in statistics (left panel) and on evaluation anxiety in statistics (right panel). The numbers are the raw regression coefficients with their 95% confidence intervals.



In a second, more detailed, model, spatial anxiety is introduced. Visuospatial working memory span is expected to first influence spatial anxiety, which, in turn, influences statistics anxiety. The results are shown in Figure 2. All variables were controlled for baseline anxiety using anxiety of asking for help. Mathematics anxiety was a mediator because it is believed that a fair amount of statistics anxiety is a carry-over of an anxiety toward mathematics, which is measurable years before statistics anxiety begins to manifest (Gibeau et al., 2023). Poor spatial skills (measured by a working memory task) is expected to be the first predictor as previous research have shown that poor spatial skills predict higher mathematics anxiety and lower mathematics performance (Ashkenazi & Velner, 2023; Delage et al., 2022; Ferguson et al., 2015; Sokolowski et al., 2019). Also, mental manipulation in spatial anxiety, used in the model, is the most relevant predictor of interpretation anxiety in statistics (Gibeau et al., 2023).

Figure 2

Model involving Working Memory Spans, Mental Manipulation Anxiety, Mathematics Anxiety predicting Interpretation Anxiety in Statistics from Study 1; the three anxiety measures and the tasks have anxiety of asking for help as covariate. The numbers are the raw regression coefficients with their 95% confidence intervals.

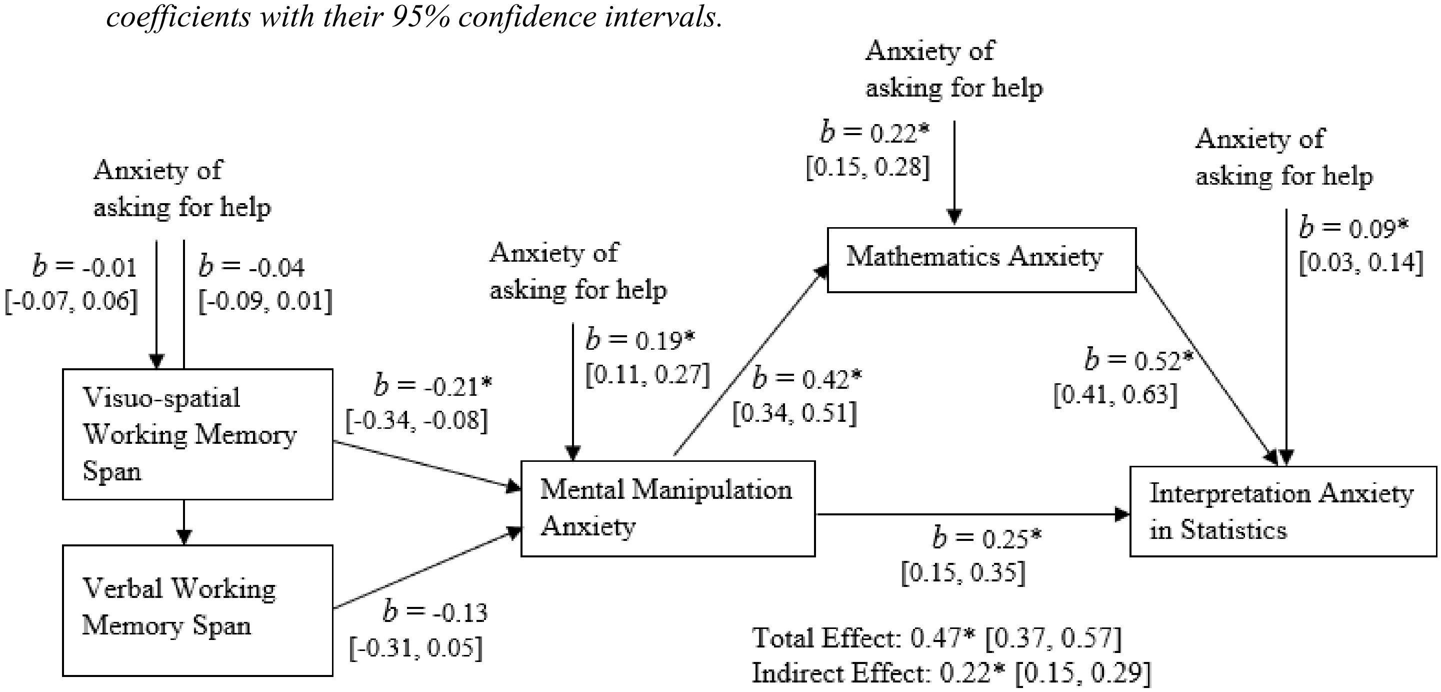


Figure 2 shows the model parameters. It fits the data well (RMSEA = .08, CFI = .97, TLI = .92, SRMR = .05). The results suggest that mental manipulation anxiety contributes to interpretation anxiety in statistics. Both working memory span tasks contribute to mental manipulation anxiety. However, the n-Back Task coefficient was not significant whereas the Corsi-Blocks Task coefficient was significant. The present model also indicates a partial mediation. The indirect path shows that mental manipulation anxiety contributes to mathematics anxiety, which contributes to interpretation anxiety in statistics, and the direct path shows that mental manipulation anxiety contributes to interpretation anxiety in statistics. Both were of similar magnitude, where the former explained 47% of the variance of interpretation anxiety in statistics. This supports the carry-over effect mentioned in the previous paragraph (also observed in Gibeau et al., 2023). However, the present study does not have a longitudinal design, consequently, no developmental conclusions can be made.

Discussion

The associations between tasks and anxiety types are not as strong as the correlations between different types of anxiety. This can be explained by the difference in data collection methodology of working memory and anxiety (i.e., tasks vs. self-reported questionnaires; Clark & Maguire, 2020, report a similar pattern). Nonetheless, the analyses support the main hypotheses. Indeed, the consistently significant associations between visuospatial working memory span, spatial anxiety and statistics anxiety suggest that spatial cognition plays a role in statistics learning among social sciences students. This is in line with previous research in numerical cognition, in which spatial skills play a role in mathematics learning through a *cognitive-emotional framework* (e.g., (Ashkenazi & Velner, 2023; Ferguson et al., 2015; Sokolowski et al., 2019). Furthermore, the mediation model also supports the influence of cognitive abilities on spatial anxiety with subsequent effect on both mathematics anxiety and statistics anxiety. This chain of associations passes through mental manipulation anxiety (with a correlation of -.19 with SAQ-M seen in Table 3) compared to navigation (r = -.14) and imagery anxiety (r = .04). This last result is also in line with previous research, in which mental manipulation anxiety is the strongest mediator in the relationship between gender and mathematics anxiety (Delage et al., 2022).

As such, visuospatial working memory span can be considered as a *resource* available to perform cognitive tasks such as learning statistics. Having this resource, however, does not mean that the person is well trained to use it. Consequently, it would be interesting to examine visuospatial *skills* (i.e., the manifested skill to visuospatially manipulate organized information). To that end, the two working memory span tasks can be replaced with ability tasks; one measuring spatial skills and one measuring verbal abilities. This is done in Study 2.

Study 2

It was observed in Study 1 that imagery anxiety had different correlations with working memory spans and mental manipulation spatial anxiety. To examine this distinction further, a measure of mental imagery vividness was added in Study 2. The reason for this exploratory analysis is the conjecture that people with more vivid images might have better spatial skills, and less spatial anxiety (but note that Blajenkova et al., 2006, argued that these two constructs may be unrelated).

Method

Participants

Study 2 was advertised to the same population as in Study 1. The only difference is that these students were enrolled in a statistics course during the 2022–2023 academic year.

Materials

First, participants provided informed consent and answered a short series of sociodemographic questions. Then, they completed three questionnaires and two tests as described below. The questionnaires, their items and the tests were presented in a randomized order. As in Study 1, all scores were rescaled on a range from 0 to 100 to ease interpretation.

Preserved Measures. The mental manipulation and the imagery subscales of the *Spatial Anxiety Questionnaire* (SAQ) were included in this study because mental manipulation and imagery should be closely related concepts and a subsidiary goal is to deepen the understanding of their relation. Similarly, the anxiety of asking for help and interpretation anxiety subscales of the *Statistical Anxiety Scale* (SAS) were used because the second is the most important aspect of statistics anxiety and the first is a good covariate to control for items bearing on anxiety.

Discarded Measures. The *Abbreviated Math Anxiety Scale* (AMAS) was not included in Study 2 as its mediation effect is now well established. Also, navigation in spatial anxiety (SAQ-N) and evaluation anxiety in statistics (SAS-E) were removed because Study 1 showed no significant correlations between each of these two variables and statistics interpretation anxiety, which is the main concern of Study 2.

New Measures. To measure spatial skills, a French translation of the *Spatial Reasoning Instrument* (SRI; original version in Ramful et al., 2017) was used. The questionnaire comprises 30 items divided into three subscales: one depicting mental manipulation (SRI-M; e.g., "Here's a picture of two dogs. Which one of the following shows the picture after a 90 degrees turn to the right?"), another depicting navigation (SRI-N; e.g., "Briana placed a hamster at the start of a maze as shown below. The hamster ran through the maze. It turned to its right, then turned left, then turned right. Where did the hamster finish?") and a third depicting imagery (SRI-I; e.g., "Jane makes a pattern by first folding a square paper into 4 parts as shown below. She then cuts the folded piece of paper into the shape shown below. Which one of the following represents the pattern obtained by unfolding the paper?"). The original authors reported good reliability (α of .85 overall) and good convergent validity (correlations between .33 and .62 with similar measures, i.e., the Card Rotation Test, Cube Comparison Test, Paper Folding Test and the Perspective Taking Test). Internal consistency ranges between .66 to .85 for all three subscales and the total scale. Note that the specific type of internal consistency used is not mentioned. Three items were removed for the present study because one was a repetition and the other two had an ambiguous wording or response.[5]

To measure verbal abilities, 15 items measuring French writing skills and vocabulary were assembled. First, five questions on verb conjugation from the QCM-Concours website (VA; http://micromagie.com/quizfaber/francais/conj2/Conjugaison2) was used. To get 1 point for each question, participants had to find the correct answer. A loosely translated example is: "We had (harm) each other without knowing," with "harmt," "harmed," "harm," "harms," and "harming" as response options. Second, 10 items measuring verbal logic were added (http://micromagie.com/quizfaber/francais/logiqueverbale/Aptitude%20verbale), which implement a general level of cognitive functioning. A loosely translated example is: "Find the word that best completes the following analogy: *For* is to *against*, what *always* is to…" with "sometimes," "only," "again," "never," and "maybe" as response options. The items were scored as correct or incorrect and the total was computed, ranging from 0 to 15, and rescaled on a range from 0 to 100 for easier interpretation. Internal consistency with the present sample of this measure is considered acceptable (see Table 5). It is also considered to have good divergent validity due to negative moderate correlations with the other measures of the present study (i.e.,

[5] The items are Item 17 (which is identical to item 8): "Briana placed a hamster at the start of a maze as shown below. The hamster ran through the maze. It turned to its right, then turned left, then turned right. Where did the hamster finish?"; Item 24: "A pedestrian crossing sign is shown below. How many lines of symmetry do the sign have?"; Item 28: "Two plans of Victoria hospital are shown below. Which building is on the corner of Aranda Street and Macquarie Street?"

SRI, SAQ, SAS, VVIQ).

To explore the level of detail of mental images, the French version of the *Vividness of Visual Imagery Questionnaire* (VVIQ-F; Campos et al., 2000; original version in Marks, 1973) was used. This is a 16-item questionnaire divided into four scenarios of four questions each (imagine a family member; imagine a sunrise; imagine a store you often go to and imagine a country scene). Participants answered on a 5-point scale: (1) “Perfectly clear and as vivid as normal vision”, (2) “Clear and reasonably vivid”, (3) “Moderately clear and vivid”, (4) “Vague and dim” and (5) “No image at all, you only ‘know’ that you are thinking of the object.” Note that the scale is actually reversed as a high score reflects a *lack of* vividness; scores ranged from 16 (perfectly clear) to 80 (no image at all) which were then rescaled on a range from 0 to 100 for easier interpretation. Previous research have reported high validity and reliability of the VVIQ (Campos, 2011). In the original article (Mars, 1973), participants were asked to complete a visualization task and the VVIQ. Participants with a high performance on the task also rated the vividness of their mental images as high. This result is an indication of good convergent validity. The questionnaire’s internal consistency with the present sample is excellent (see Table 5). Good divergent validity is indicated by the lack of correlation between the VVIQ and the other measures used herein, except with the SAQ, which is to be expected (see Table 6).

Finally, one more item was added to the questionnaire. The item asked participants to rate their general anxiety level from 1 to 10. The item was added to provide a baseline anxiety level which will be used as a control alongside anxiety of asking for help in statistics (see Allen et al., 2022, for the use of single-item measurements). This measure was inspired by the Single-Item Math Anxiety questionnaire (Núñez-Peña et al., 2014).

Procedure

The procedure is the same as in Study 1.

Results

Data screening

A total of 679 students clicked on the link to the questionnaire provided by their instructor and 374 completed it. This represents a completion rate of 55.1%; better than Study 1 during the COVID-19 pandemic (31.5%) but lower than past studies conducted before the first lockdown.

There were 100 missing responses (about 0.3% of the items) and no outlier; all the scales and subscales had responses. The missing items' responses were imputed with the mean of the other responses. The distributions appear relatively normal, except for the SRI which has a small asymmetry in the scores (skewness = -0.67). It is possible that some of the items were too easy, resulting in a ceiling effect. The distribution of VA and of the VVIQ were also skewed (-1.44 and 1.15, respectively).

Descriptive Statistics

The descriptive statistics are reported in Table 5. The majority of participants selected the age range 18-24 years-old (86.4%) which is also the lower age range selected. The higher age range selected by one participant is 55-64 years-old. Likewise, women composed 86.4% of the sample; men composed 11.2% and 2.4% self-identified as other genders. Participants were from North America (10.2%) and Europe (89.8%).

All the Cronbach's alphas and McDonald's omegas were good for SAS (α for SAS-A: .97, SAS-I: .88 and over all: .93; ω for SAS-A: .97, SAS- I: .88 and over all: .96), replicating the ones reported in the original publications. The Cronbach's alpha and McDonald's omega regarding the Verbal ability items is weak (α = .44; ω = .46). Of those, two conjugation items had poor performance (14% and 19% correct respectively for the 11th and 14th items); when removed, α rose to .47 and ω rose to .50. It was decided to keep these two items. Despite the low reliability of this measure, it is kept herein for transparency.

Table 5

Descriptive Statistics of all Variables in Study 2

Variables	*N*	Mean [95% CI]	SD	Min.	Max.	Skewness	Cronbach's α	McDonald's ω
VA	374	73.1 [65.6, 80.6]	11.4	20	100	-1.09	.44	.46
GenAnx	374	63.6 [61.2, 66.1]	23.8	10	100	-0.53	NA	NA
SRI	374	74.0 [72.2, 75.3]	17.2	22	100	-0.67	M : .70; N : .55; I : .63; all : .81	M : .70; N : .57; I : .63; all : .83
SAQ	374	39.7 [37.8, 41.6]	18.7	0	100	0.06	M: .89; I : .85; all : .89	M : .89; I : .85; all : .91
SAS	374	46.5 [44.1, 48.8]	23.1	0	100	-0.04	A : .97, I : .88; all : .93	A : .97; I : .88; all : .96
VVIQ	374	42.7 [41.2, 44.3]	15.3	19	100	1.15	.92	.94

Notes. VA signifies *Verbal Abilities*; GenAnx signifies the question about participants' general anxiety level; SRI signifies *Spatial Reasoning Instrument*; SAQ signifies the *Spatial Anxiety Questionnaire*; SAS signifies the *Statistics Anxiety Scale*; VVIQ signifies *Vividness of Visual Imagery Questionnaire* (where high scores mean a lack of vividness). All the questionnaires and tasks are reported on a scale of 0 to 100 points.

Correlations between Questionnaires' Total Scores and Tests

Table 6 shows the correlation matrix. Spatial anxiety and statistics anxiety were strongly correlated (r = .48), which replicates the results of Study 1 and the results of a previous study (Gibeau et al., 2023). The spatial skills test (SRI) was negatively and weakly associated to both anxiety types (SAQ: r = -.18; SAS: r = -.16). Despite the weak correlations, this is congruent with the main hypotheses. Indeed, social sciences students with higher spatial skills would tend to be less spatially anxious and less anxious with respect to statistics, which supports the claim that spatial skills play a role in understanding statistics.

Table 6

Correlation Matrix of all Questionnaires' total score in Study 2 (N = 374)

Variables	GenAnx	VA	SRI	SAQ	SAS
VA	-.08				
	[-.18, .02]				
SRI	-.13	.38			
	[-.23, -.03]	[.29, .47]			
SAQ	.26	-.12	-.18		
	[.17, .35]	[-.22, -.02]	[-.28, -.08]		
SAS	.44	-.14	-.16	.48	
	[.35, .51]	[-.24, -.04]	[-.26, -.06]	[.40, .55]	
VVIQ	-.04	-.05	-.08	.13	.05
	[-.14, .06]	[-.16, .05]	[-.18, .02]	[.03, .23]	[-.05, .15]

Notes. VA signifies *Verbal Abilities*; SRI signifies *Spatial Reasoning Instrument*; SAQ signifies the *Spatial Anxiety Questionnaire* (*mental manipulation* and *imagery* only); SAS signifies the *Statistics Anxiety Scale* (*interpretation* and *asking for help* only); VVIQ signifies *Vividness of Visual Imagery Questionnaire* (where high scores mean a lack of vividness); GenAnx signifies the question about participants' general anxiety level. The numbers between brackets are the 95% confidence intervals of the correlation coefficient

Verbal abilities are weakly and negatively correlated to spatial anxiety and statistics anxiety (SAQ: $r = -.12$; SAS: $r = -.13$). The relations are similar to the one between SRI and these forms of anxiety but are weaker. Finally, SRI and VA are moderately correlated ($r = .38$), suggesting some shared variance between the two measures.

Maybe surprisingly at first, lack of imagery vividness was weakly related to spatial skills; the correlation between VVIQ and SRI is not different from zero ($r = -.08$). However, according to Delem et al. (2025), it is not surprising. The results of their study suggest that aphantasics have intact spatial representations. The VVIQ is weakly and negatively correlated with verbal abilities (also $r = -.08$). As mentioned above, this latter measure had a weak Cronbach's alpha, consequently, no strong claims can be made but this absence of relation with verbal abilities would be congruent with expectations.

Finally, general anxiety moderately and positively correlates with spatial and statistics anxiety (r = .26 and .44, respectively; both CIs do not include zero). This variable also significantly correlates with spatial skills but negatively instead of positively (r = -.13). The other correlations involving general anxiety were not significant as evidenced by the confidence intervals that include zero. These moderate correlations support the conjecture from Study 1 that anxiety questionnaires have a common component. In the previous study, general anxiety was controlled for by using anxiety of asking for help in statistics; here it will be more directly controlled by the general anxiety variable.

Correlations between Subscales and Tests

Table 7 shows the complete correlation matrix. The results replicate again previous research where Mental Manipulation Anxiety of the SAQ correlates more strongly with Interpretation Anxiety of the SAS (r = .52) than with the Anxiety of Asking for Help of the SAS (r = .30; when comparing these rs, z = -3.63 and p = .0003). The SAQ-M subscale correlated well with Mental Manipulation skills –compared with Navigation skills– of the SRI (r = -.33 compared to r = -.19; z = -2.05 and p = .04). This was expected as both SAQ-M and SRI-M measure a similar construct, i.e., manipulation of elements. Statistics interpretation anxiety was negatively related to SRI-M and to SRI-I (r = -.25 and -.22, respectively) but less with SRI-N (r = .09). However, all three scales of the SRI were strongly intercorrelated (all rs > .47), which indicates that they measure very closely related concepts.

Table 7

Correlation Matrix of Subscales in Study 2

Variables	GenAnx	VA	SRI-M	SRI-N	SRI-I	SAQ-M	SAQ-I	SAS-A	SAS-I
VA	-.08 [-.18, .02]								
SRI-M	-.08 [-.18, .02]	.32 [.23, .41]							
SRI-N	-.12 [-.22, -.02]	.35 [.26, .44]	.50 [.42, .57]						
SRI-I	-.13 [-.23, -.03]	.31 [.21, .40]	.62 [.55, .67]	.47 [.39, .54]					
SAQ-M	.25 [.15, .34]	-.11 [-.21, -.01]	-.33 [-.41, -.23]	-.19 [-.29, -.09]	-.27 [-.36, -.17]				
SAQ-I	.19 [.09, .29]	-.09 [-.19, .01]	.05 [-.05, .15]	-.04 [-.14, .06]	.04 [-.06, .14]	.43 [.35, .51]			
SAS-A	.41 [.32, .49]	-.08 [-.18, .02]	-.05 [-.15, .05]	-.04 [-.14, .06]	-.04 [-.14, .06]	.30 [.20, .39]	.30 [.20, .39]		
SAS-I	.30 [.21, .39]	-.16 [-.26, -.06]	-.25 [-.34, -.15]	-.09 [-.19, .01]	-.22 [-.31, -.12]	.52 [.44, .59]	.26 [.17, .36]	.39 [.30, .47]	
VVIQ	-.04 [-.14, .06]	-.05 [-.16, .05]	-.03 [-.13, .07]	-.13 [-.23, -.03]	-.07 [-.17, .03]	.08 [-.03, .18]	.15 [.05, .25]	.08 [-.02, .18]	-.02 [-.12, .08]

Notes. VA is the conjugation questions of the *Verbal Ability Quiz*; GenAnx signifies the question about participants' general anxiety level; SRI-M is the Mental Manipulation of the *Spatial Reasoning Instrument*; SRI-N is the Spatial Navigation subscale of the *Spatial Reasoning Instrument*; SRI-I is the Imagery subscale of the *Spatial Reasoning Instrument*; SAQ-M is the Mental Manipulation subscale of the *Spatial Anxiety Questionnaire*; SAQ-I is the Imagery subscale of the *Spatial Anxiety Questionnaire*; SAS-A is the Anxiety of Asking for Help subscale of the *Statistics Anxiety Scale*; SAS-I is the Interpretation subscale of the *Statistics Anxiety Scale*; Analogy is the verbal logic questions of the *Verbal Ability Quiz*; VVIQ is the *Vividness of Visual Imagery Questionnaire*, where a high score means a lack of imagery vividness. The numbers between brackets are the 95% confidence intervals of the correlation coefficient

Verbal abilities are positively related to all three subscales of spatial skills (with *r*s between .35 and .40). Moderate correlations could be expected as (a) the SRI has a language dimension since, to have the correct answer, one must understand the written question; (b) vocabulary is an indirect measure of IQ and therefore, high-performing participants on verbal abilities may likewise tend to be high functioning in other areas as well (e.g., Laasonen et al., 2020).

The correlations between Imagery Anxiety and Manipulation anxiety in the SAQ and Anxiety of Asking for Help in the SAS were fairly high (both $r = .30$, as in Study 1 where they were .26 and .29, respectively), which suggests that the imagery and manipulation subscales of the SRI would also be related to SAS-A. Surprisingly, they were not ($r = -.05$ and $r = -.04$, respectively). This strengthens the suggestion that the association between anxiety of asking for help (which is more social in nature), and imagery anxiety is solely a general anxiety relation and not a manifestation of a lack of abilities. This is also indicated by the correlations involving general anxiety. Indeed, general anxiety was positively correlated with all anxiety subscales with a mean *r* of .29. On the other hand, general anxiety was negatively correlated with navigation skills and imagery skills with an *r* of -.12 and -.13 respectively. This is why, in Study 1, anxiety of asking for help was used in the computation of partial correlations and as a covariate in the models. This is done again hereafter with the addition of general anxiety.

Concerning the VVIQ, it can be observed that it is weakly and positively related to the imagery subscale of the SAQ ($r = .15$). This suggests that people with higher imagery anxiety lack vivid images, and vice versa. Considering these results, one would expect to observe correlations between the VVIQ and the imagery subscale of the SRI. This is not what was observed. Indeed, imagery vividness is not related to the imagery subscale of the SRI (and neither it is to SRI-M). However, the second-highest correlation involving VVIQ is with the navigation subscale of the SRI. This is surprising at first but when looking at the items of the VVIQ, one can understand why that is; three out of four scenarios concern a physical place (e.g., a store one often *goes* to).

Table 8 shows the regular correlations and the semi-partial correlations of general anxiety, VA, all subscales of the SRI with regards to interpretation anxiety in statistics. As seen, the only skills having moderate unique variances that predict interpretation anxiety in statistics are mental manipulation skills (partial correlations of -.15). Both Imagery and Navigation skills (SRI-I and SRI-N) dropped to partial correlations of .09 and -.06 respectively. In addition, general anxiety (GenAnx) remained a good predictor when the other variables were partial out, reinforcing the idea that Interpretation of statistics anxiety has an anxiety component instead of only being a lack of skills.

Table 8

Pairwise and Semi-Partial Correlations for five Predictors of Statistics Interpretation Anxiety (SAS-I) in Study 2

	Correlations	
Variables	Pairwise	Semi-Partial
GenAnx	.30	.28
VA	-.16	-.08
SRI-M	-.25	-.15
SRI-N	-.09	.09
SRI-I	-.22	-.06

Note. The pairwise correlations are reproduced from Table 7 for easier comparison.

Path Regression Models

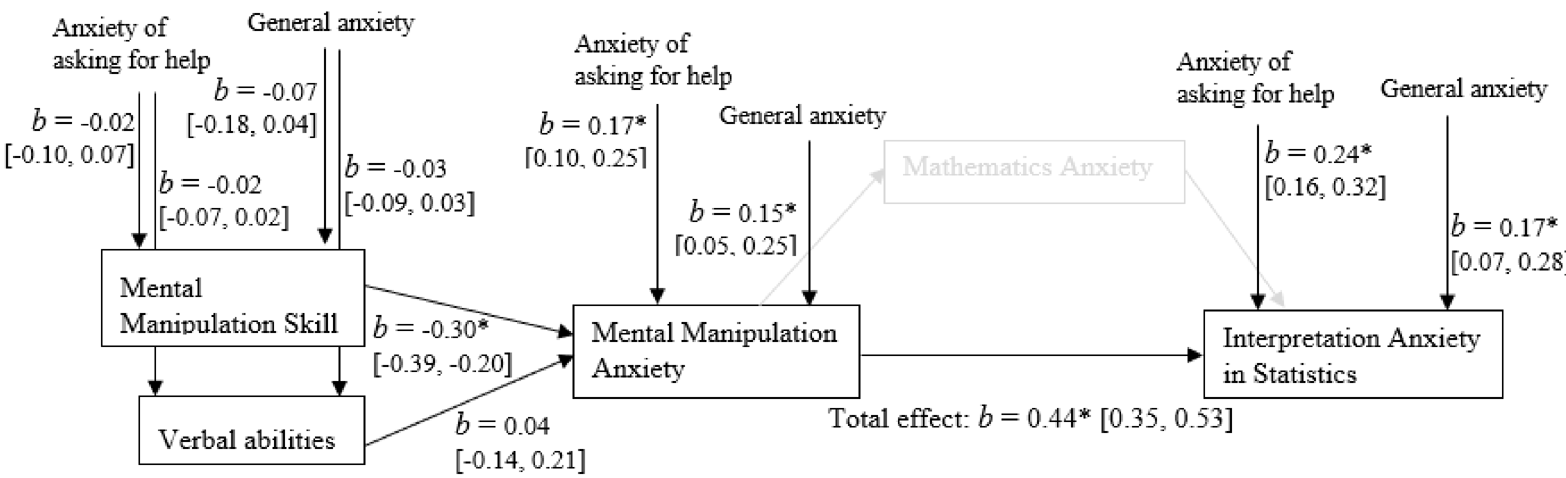
A model was estimated using the mediation command in the MPlus software (version 7) to test the hypothesis that mental manipulation skills, compared to verbal abilities, predict mental manipulation anxiety, which in turn, predicts statistics interpretation anxiety. This model is analogous to the model of Study 1 (seen in Figure 2) and has the same rationale, but with skills replacing spans, removing Mathematics Anxiety as it is not part of the current study, and replacing gender with general anxiety as the second covariable.

Figure 3 shows the resulting model. As can be seen, the model fits the data well (RMSEA = .09, CFI = .96, TLI = .89, and SRMR = .05). The RMSEA is higher than the .05 recommendation. However, this fit index tend to be inaccurate in models with a large number of observed variables (Shi et al., 2020). The relation between SAQ-M and SAS-I (unstandardized regression coefficient of 0.44) has a similar magnitude as the total effect in the previous study, which indicates that this finding is replicated. The coefficient from the mental manipulation skills to mental manipulation anxiety is moderate and negative: those with better skills in this spatial dimension experience lower spatial anxiety, or vice versa. Verbal abilities, on the other hand, have no noticeable effect. In this model, the covariates *anxiety of asking for help* and *general anxiety* were used to control the various predictors and dependent variable prior to fit the whole model.

The model was replicated once with navigation skills and a second time with imagery skills, instead of mental manipulation skills. These changes barely had any effect on the goodness of fit and predicted mental manipulation anxiety with a little smaller but quite similar coefficients. Considering the small differences, these two models are not presented herein.

Figure 3

Mediation Model Between Spatial Skills, Mental Manipulation Anxiety, and Interpretation Anxiety in Statistics from Study 2. The numbers are the raw regression coefficients with their 95% confidence intervals. Mathematics anxiety is greyed as it was not measured in Study 2



Discussion

The primary objective of Study 2 was to further investigate the findings from Study 1 by replacing memory span tasks with skills tests to assess their association with spatial problem solving skills, statistics anxiety, and spatial anxiety, compared to verbal abilities. The results of Study 2 strongly support the hypothesis that better visuospatial skills are inversely related to mental manipulation anxiety and statistics interpretation anxiety.

The negative correlation between mental manipulation skills and spatial anxiety, as well as statistics interpretation anxiety, suggests that social sciences students with enhanced visuospatial skills are less likely to experience anxiety in these areas, or vice versa. This finding aligns with previous research indicating the importance of spatial skills in reducing academic anxieties (Cantinotti et al., 2017; Vigil-Colet et al., 2008). In contrast, verbal abilities showed negligible association with mental manipulation anxiety, which emphasizes the domain-specific nature of the anxiety experienced in statistics interpretation. Interestingly, the measure of mental imagery vividness, introduced as an exploratory variable, did not correlate significantly with statistics interpretation anxiety. This result can be understood by considering the nature of interpretation anxiety, which is more about manipulating data mentally rather than visualizing it in detail (Cantinotti et al., 2017). Thus, being able to conjure up detailed imagery might not be as critical as the ability to mentally manipulate abstract statistical concepts (see Delem et al., 2025, for results supporting this hypothesis). Considering that the VVIQ is solely visual, the results might have been different – or the distinction between imagery and manipulation could have been seen – if the *Object-Spatial Imagery and Verbal Questionnaire* (OSIVQ; Blazhenkova & Kozhevnikov, 2009) was used instead. This measure may be an interesting addition in future research.

The path regression model supported the hypothesis that mental manipulation skills predict mental manipulation anxiety, which in turn predicts statistics interpretation anxiety. This chain of associations underscores the critical role of spatial skills in academic contexts involving statistical interpretation. Moreover, these findings suggest that interventions aimed at improving spatial skills could potentially reduce statistics anxiety, thereby facilitating better learning outcomes in statistics courses. The results also reinforce the idea that while general anxiety contributes to various forms of academic anxiety, specific skills and domain-specific anxieties play crucial roles in the learning process. The consistency of these findings with prior research (Gibeau et al., 2023) strengthens the validity of the proposed model and highlights the importance of considering both cognitive abilities and anxiety management in educational strategies.

Overall, Study 2 provides robust evidence for the significant impact of spatial skills on anxiety in statistical interpretation, suggesting practical applications for educational practices and interventions aimed at improving student performance and reducing anxiety in statistics learning contexts.

General Discussion

Today's society navigates within a Big Data era. From private to public sectors, from healthcare to social media companies, massive amounts of data are being collected and stored within the digital space (McMahon et al., 2020; Villarejo-Ramos et al., 2021). Despite the necessity for statistical literacy, many individuals struggle with interpreting statistical information due to inherent difficulties in understanding probabilistic, quantitative data, and graphical representations (Béland et al., 2017; Husereau et al., 2024; Zang & Cousineau, 2025).

When learning about statistics, most students have high levels of anxiety, making it difficult to understand the subject (Cousineau & Harding, 2017; Garfield & Ben-Zvi, 2008; Gibeau et al., 2023; Onwuegbuzie, 2004; Onwuegbuzie & Wilson, 2003). The goal of the two studies presented herein was to identify predictors of statistics anxiety to eventually be able to decrease its negative effects on students and promote an increased understanding of the subject.

The results of the two studies presented in this text reveal patterns of correlations involving students' anxiety toward interpreting statistical information. In Study 1, visuospatial working memory span influences the level of mental manipulation anxiety, which in turns influences the level of statistics interpretation anxiety (directly and indirectly through mathematics anxiety). This suggests that social sciences students with a lower capacity to retain visuospatial information also have higher anxiety toward mentally manipulating elements, higher anxiety toward mathematics, and higher anxiety toward interpreting statistical information, or vice versa. In Study 2, a similar pathway was identified when memory spans were replaced by abilities. Taken together, these studies highlight the influence of spatial processing in statistics anxiety. The experimental design does not allow inferring a precursory role to spatial processing nor a causal role. Both mathematics anxiety and spatial anxiety are observed in children as young as 6-8 years of age, (Gunderson et al., 2013; Ramirez et al., 2016). Given that children at this age have not yet been exposed to formal statistics in school, it seems unlikely that exposure to statistics can have caused their math and spatial-related anxieties. Rather, it seems more likely that math and spatial anxiety, in general, comes online before statistics anxiety. However, this is only a hypothesis.

One difference between the two studies, presented herein, lies in the magnitude of the effects. In Study 1, the lag between the strength of association between memory spans and mental manipulation anxiety (-.21 for the Corsi vs. -.13 for the n-Back) is nearly three times smaller than the corresponding lag in Study 2 (-.30 for the SRI-M vs. .04 for the VA). One possible reason is that the n-Back Task also measures central executive functions, which are necessarily present in the Corsi-Blocks Task. Hence, the n-Back Task might lack specificity, leading to a conservative estimate of the lag. Another reason is that the verbal ability test that was assembled had poor psychometric properties (a Cronbach α of .44 over 15 items is quite low), leading to a liberal estimate of the lag. Better-controlled measures are recommended for future research.

Further, the pattern of results of the VVIQ was intriguing. It correlated moderately with navigation skills, perhaps because the items were biased toward depicting traveling to an area (and navigation is sometimes performed by retrieving a sequence of landmarks; Epstein et al., 2017). It also correlates moderately with imagery anxiety. Yet, both imagery anxiety and navigation abilities were uncorrelated. Finally, VVIQ was not correlated to interpretation of statistics anxiety. Such results could be reconciled by the hypothesis that visual skills and spatial skills may be distinct sets of skills.

Implications for Statistics Education

It is known that anxiety negatively impacts working memory, reducing the quantity of relevant information stored and the quality of its processing (Dutke & Stöber, 2001; Eysenck & Calvo, 1992). When performing a task, it is necessary to focus on relevant information and to disregard distractors. People experiencing anxiety may pay less attention to relevant information, have intrusive thoughts and show impaired spatial processes and reduced working memory spans, all of which reduces the amount of cognitive resources one has available. Consequently, they are less likely to complete the task successfully and therefore, impeding the quality of learning (Berggren & Derakshan, 2013; Lisica et al., 2022).

As observed in many studies, performance in statistics courses is predicted by anxiety toward the subject (Cantinotti et al., 2017; Hoegler & Nelson, 2018; Macher et al., 2015; McIntee et al., 2022; however, see Mackinnon et al., 2025 for an opposite result). To improve students' performance in statistics, it is crucial to address a predictor: statistics anxiety. The findings from this research suggest a potential new set of predictors that could help in reducing statistics anxiety. The first step involves mental manipulation anxiety that could be reduced by improving mental manipulation skills through spatial training. It is also possible that simple exposure to mental manipulation activities without avoidance in a calm and safe environment can reduce the anxiety, similar to clinical treatments. The second step focuses on statistics interpretation anxiety that could be decreased by addressing mental manipulation anxiety, as this type of anxiety predicts levels of anxiety related to interpreting statistical information. Reducing mental manipulation anxiety could influence the levels of statistics interpretation anxiety. Again, it is possible that exposure to interpretation tasks without avoidance in a calm and safe environment can reduce the anxiety. This would be especially pertinent to examine considering that these types of tasks are often required in a class or exam contexts, which may not be clam and safe environments.

While the ability to retain visuospatial information may not be easily enhanced, there is evidence supporting the efficacy of cognitive training in improving mental manipulation skills (Diamond & Ling, 2019). It is acknowledged that the present study's correlational design does not clarify whether improving visuospatial working memory directly helps reduce mental manipulation anxiety. However, mental manipulation skills can be more readily improved with practice through targeted activities and interventions, as commonly done with children (e.g., Burte et al., 2017). These studies suggest that incorporating spatial activities into the curriculum could help improve students' spatial skills, which predict widespread apprehension toward statistics. Implementing spatial training exercises and dynamic visualization techniques within statistics education could provide a dual benefit: enhancing cognitive skills relevant to statistical tasks and reducing anxiety. Educators could integrate these activities into their teaching strategies, creating a more supportive learning environment that addresses both cognitive and emotional challenges faced by students in statistics courses. This pedagogical approach to improve performance in statistics needs formal testing as it is based on the present correlational studies.

Limitations

The results presented herein are positive and congruent with a contributing relation of spatial cognition and statistics understanding. There are a few caveats that are acknowledged here. First, this is a correlational study. It is hypothesized that there may be a cascade of influences from reduced spatial skills and lower visuospatial working memory to higher statistics anxiety. However, it is impossible to infer the causality of this conclusion with the present data. Also, statistics anxiety probably has a historical aspect: learners who experienced difficulties understanding preliminary concepts useful to statistics in their past possibly developed heightened levels of statistics anxiety. To examine this developmental hypothesis, a longitudinal study, where measurement begins in early high school (or even earlier), is required. The results of the present study indicate that statistics anxiety has a few spatial predictors, but this does not mean that these predictors cause a heightened anxiety toward statistics.

Also, the sample is composed only of students in social sciences and the majority are women, which may have influenced the results and conclusions. These programs are not mathematically-oriented. Presumably, many students likely enrolled in these programs because they do not perceive themselves as math- or science-oriented (Esnard et al., 2021; Paechter et al., 2017). Many possibly assumed that there would be no mathematics or statistics courses when enrolling in such programs. The element of surprise and the fact that they do not perceive themselves as number-oriented people may exacerbate the amount of anxiety in such university students, in a form of stereotype effect. The fact that women are also typically more anxious toward statistics may also be explained in part by stereotype effects. However, some studies have shown that STEM students may also have a high level of anxiety toward statistics or mathematics (e.g., Daker et al., 2021). This suggests a widespread discomfort toward these two subjects whether or not people are number- or science-oriented.

Finally, some measures used herein have limitations. Some argue that the *Corsi-Blocks Task* does not fully capture visuospatial working memory but short-term memory instead. The *Spatial Reasoning Instrument* seems to have a few items with a ceiling effect. The distribution of this measure is somewhat skewed where there are more participants at the high end of the distribution. Similarly, participants seem to be very good in the n-Back Task used in Study 1. Most participants finished all conditions without reaching the maximum number of errors allowed. Consequently, the 3-back condition may be too easy for the studied population. Another measure that produces skewed scores, but expected to be, is the *Vividness of Visual Imagery Questionnaire*. More participants reported high levels of vividness, as reported in previous studies (e.g., Dance et al., 2022). Also, the measures were presented online, which may have tapped into participants' spatial skills and anxiety. If the measures were done paper-pencil, participants may not have reported such high spatial anxiety nor solicited their visuospatial working memory as much. On a side note, because memory is impacted by age, and that there are a few participants in their 40s and 50s, this variable should have been controlled for in the path regression analyses. Their age could have influenced the results.

Future research should examine which spatial activities help students develop their spatial skills. In doing so, these activities could be implemented in statistics courses to facilitate students' learning, understanding and performance through learning how to use their spatial skills in statistics. This could further extend the literature on statistics education by proposing applicable solutions to instructors.

Take-Class Message

The message to keep in mind is that incorporating spatial activities and spatial visualizations into the teaching of statistics may be beneficial for students as it is in spatial intervention studies (e.g., Lowrie & Logan, 2023). By asking students to practice their spatial skills and to teach abstract concepts through visual stimuli instead of formulas and definitions, it may improve their spatial skills, reduce their anxiety and facilitate their understanding by providing concrete elements. A formal test of this hypothesis is yet to be performed.

As many statistics courses have two components (lectures and laboratories), it should be relatively easy to incorporate spatial training within the class. For example, the first few laboratories can ask students to participate in spatial tasks, like Tetris or origami folding (Burte et al., 2017). Then, the data collected with these tasks can be further used in demonstrating how to perform analyses, as it is often recommended to use real data when teaching statistics (e.g., Carver et al., 2016).

References

Abdi, H. (2007) Part (semi partial) and partial regression coefficients. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 1-9). Sage. https://personal.utdallas.edu/~herve/Abdi-PartialRegressionCoefficient2007-pretty.pdf

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science: A call to action. *European journal of psychological assessment*. *38*(1), 1-5. https://doi.org/10.1027/1015-5759/a000699

Ashkenazi, S., & Velner, H. (2023). The interplay between math performances, spatial abilities, and affective factors: The role of task. *Trends in Neuroscience and Education*, *33*, 1-9. https://doi.org/10.1016/j.tine.2023.100211

Atit, K., Power, J. R., Pigott, T., Lee, J., Geer, E. A., Uttal, D. H., Ganley, C. M., & Sorby, S. A. (2022). Examining the relations between spatial skills and mathematical performance: A meta-analysis. *Psychonomic Bulletin & Review*, *29*(3), 699-720. https://doi.org/10.3758/s13423-021-02012-w

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), 136-140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556–559. https://www.jstor.org/stable/2876819

Baloğlu, M. (1999). A Comparison of mathematics anxiety and statistics anxiety in relation to general anxiety. *ERIC*, 1–31. https://eric.ed.gov/?id=ED436703

Baloğlu, M. (2004). Statistics anxiety and mathematics anxiety: Some interesting differences I. *Educational Research Quarterly*, *27*(3), 38-48. https://eric.ed.gov/?id=EJ792861

Béland, S., Cousineau, D., & Loye, N. (2017). Utiliser le coefficient omega de McDonald à la place de l'alpha de Cronbach. *McGill Journal of Education*, *52*(3), 791-804. https://doi.org/10.7202/1050915ar

Bennett-Pierre, G., & Gunderson, E. A. (2023). Fiber arts require spatial skills: How a stereotypically feminine practice can help us understand spatial skills and improve spatial learning. *Sex Roles*, *88*(1), 1-16. https://doi.org/10.1007/s11199-022-01340-y

Berggren, N., & Derakshan, N. (2013). Attentional control deficits in trait anxiety: Why you see them and why you don't. *Biological Psychology*, *92*(3), 440-446. https://doi.org/10.1016/j.biopsycho.2012.03.007

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-Spatial Imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, *20*(2), 239-263. https://doi.org/10.1002/acp.1182

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*(5), 638-663. https://doi.org/10.1002/acp.1473

Bonnot, V., & Croizet, J.-C. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of Experimental Social Psychology*, *43*(6), 857-866. https://doi.org/10.1016/j.jesp.2006.10.006

Bradley-Garcia, M., & Bolton, V. (2023). Programming an n-Back Task in Qualtrics using HTML and JavaScript. *The Quantitative Methods for Psychology*, *19*(2), 136-155. https://doi.org/10.20982/tqmp.19.2.p136

Burte, H., Gardony, A. L., Hutton, A., & Taylor, H. A. (2017). Think3d!: Improving mathematics learning through embodied spatial training. *Cognitive Research: Principles and Implications*, *2*, 1-8. https://doi.org/10.1186/s41235-017-0052-9

Campos, A. (2011). Internal consistency and construct validity of two versions of the revised vividness of visual imagery questionnaire. *Perceptual and Motor Skills*, *113*(2), 454-460. https://doi.org/10.2466/04.22.PMS.113.5.454-460

Campos, A., Chiva, M., & Moreau, M. (2000). Alexithymia and mental imagery. *Personality and Individual Differences*, *29*, 787-791. https://doi.org/10.1016/S0191-8869(99)00231-7

Cantinotti, M., Lalande, D., Ferlatte, M.-A., & Cousineau, D. (2017). Validation de la version francophone du Questionnaire d'anxiété statistique (SAS-F-24). *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *49*(2), 133-142. https://doi.org/10.1037/cbs0000074

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Roswell, G. H., Velleman, P., Witmer, J., & Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report 2016*. American Statistical Association. https://commons.erau.edu/publication/1083

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal*, *9*(1), Article 1. https://doi.org/10.52041/serj.v9i1.385

Chiou, C.-C., Wang, Y.-M., & Lee, L.-T. (2014). Reducing statistics anxiety and enhancing statistics learning achievement: Effectiveness of a one-minute strategy. *Psychological Reports*, *115*(1), 297-310. https://doi.org/10.2466/11.04.PR0.115c12z3

Clark, I. A., & Maguire, E. A. (2020). Do questionnaires reflect their purported cognitive functions? *Cognition*, *195*, Article 104114. https://doi.org/10.1016/j.cognition.2019.104114

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences* (pp. xxi, 490). Lawrence Erlbaum.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* [Doctoral dissertation, McGill University]. https://escholarship.mcgill.ca/concern/theses/05741s554

Cousineau, D. (2020). How many decimals? Rounding descriptive and inferential statistics based on measurement precision. *Journal of Mathematical Psychology*, *97*, 1-8. https://doi.org/10.1016/j.jmp.2020.102362

Cousineau, D., & Harding, B. (2017). Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions. *Revue de psychoéducation, 46*(2), 397-419. https://doi.org/10.7202/1042257ar

Cui, S., Zhang, J., Guan, D., Zhao, X., & Si, J. (2019). Antecedents of statistics anxiety: An integrated account. *Personality and Individual Differences*, *144*, 79-87. https://doi.org/10.1016/j.paid.2019.02.036

Daker, R. J., Gattas, S. U., Sokolowski, H. M., Green, A. E., & Lyons, I. M. (2021). First-year students' math anxiety predicts stem avoidance and underperformance throughout university, independently of math ability. *NPJ Science of Learning*, *6*(1), 1-13. https://doi.org/10.1038/s41539-021-00095-7

Dance, C. J., Ipser, A., & Simner, J. (2022). The prevalence of aphantasia (imagery weakness) in the general population. *Consciousness and Cognition*, *97*, Article 103243. https://doi.org/10.1016/j.concog.2021.103243

Davis, H. E., Stack, J., & Cashdan, E. (2021). Cultural change reduces gender differences in mobility and spatial ability among seminomadic pastoralist-forager children in northern Namibia. *Human Nature*, *32*(1), 178-206. https://doi.org/10.1007/s12110-021-09388-7

Delage, V., Trudel, G., Retanal, F., & Maloney, E. A. (2022). Spatial anxiety and spatial ability: Mediators of gender differences in math anxiety. *Journal of Experimental Psychology:*

General, *151*(4), 921-933. http://dx.doi.org/10.1037/xge0000884

Delem, M., Turkben, S., Cavalli, E., Cousineau, D., & Plancher, G. (2025). Unsupervised clustering reveals spatial and verbal cognitive profiles in aphantasia and typical imagery. *Neuropsychologia*, *219*, Article 109279. https://doi.org/10.1016/j.neuropsychologia.2025.109279

de Paula, J. J., Malloy-Diniz, L. F., & Romano-Silva, M. A. (2016). Reliability of working memory assessment in neurocognitive disorders: A study of the Digit Span and Corsi Block-Tapping tasks. *Revista Brasileira de Psiquiatria, 38*(3), 262-263. https://doi.org/10.1590/1516-4446-2015-1879

Dimond, A., & Ling, D. (2019). Review of the evidence on, and fundamental questions about, efforts to improve executive functions, including working memory. In J. M. Novik, M. F. Bunting, M. R. Dougherty, & R. W. Engle (Eds.), *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development* (pp. 143-431). Oxford University Press. https://doi.org/10.1093/oso/9780199974467.003.0008

Dutke, S., & Stöber, J. (2001). Test anxiety, working memory, and cognitive performance: Supportive effects of sequential demands. *Cognition & Emotion*, *15*(3), 381-389. https://doi.org/10.1080/02699930125922

Edirisooriya, M., & Lipscomb, T. (2021). Gender influence on statistics anxiety among graduate students. *Journal of Research in Science Mathematics and Technology Education*, *4*(1), 223-234. https://doi.org/10.31756/jrsmte.421

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*(11), 1504-1513. https://doi.org/10.1038/nn.4656

Esnard, T. R., Alladin, F. M., & Samlal, K. C. (2021). Prior mathematics performance, statistics anxiety, self-efficacy and expectations for performance in statistics: A survey of social sciences students in a Caribbean institution of higher education. *Statistics Education Research Journal*, *20*(1), 1-22. https://doi.org/10.52041/serj.v20i1.98

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, *6*(6), 409-434. https://doi.org/10.1080/02699939208409696

Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences*, *39*, 1-12. https://doi.org/10.1016/j.lindif.2015.02.007

Fullerton, J. A., & Umphrey, D. (2016). Statistics anxiety and math aversion among advertising students. *Journal of Advertising Education*, *20*(1–2), 135-143. https://doi.org/10.1177/10980482160201-216

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting research and teaching practice*. Springer.

Geary, D. C., Hoard, M. K., & Nugent, L. (2023). Boys' advantage in solving algebra word problems is mediated by spatial abilities and mathematics anxiety. *Developmental Psychology*, *59*(3), 413-430. https://doi.org/10.1037/dev0001450

Geary, D. C., Hoard, M. K., Nugent, L., & Scofield, J. E. (2021). In-class attention, spatial ability, and mathematics anxiety predict across-grade gains in adolescents' mathematics achievement. *Journal of Educational Psychology*, *113*(4), 754-769. https://doi.org/10.1037/edu0000487

Gibeau, R.-M. (2021). The Corsi Blocks task: Variations and coding with jsPsych. *The Quantitative Methods for Psychology*, *17*(3), 299-311.

https://doi.org/10.20982/tqmp.17.3.p299

Gibeau, R.-M., Maloney, E. A., Béland, S., Lalande, D., Cantinotti, M., Williot, A., Chanquoy, L., Simon, J., Boislard-Pépin, M.-A., & Cousineau, D. (2023). The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender. *Journal of Numerical Cognition*, *9*(1), 16-43. https://doi.org/10.5964/jnc.8199

Giofrè, D., Donolato, E., & Mammarella, I. C. (2018). The differential role of verbal and visuospatial working memory in mathematics and reading. *Trends in Neuroscience and Education*, *12*, 1-6. https://doi.org/10.1016/j.tine.2018.07.001

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2013). Teachers' spatial anxiety relates to 1st- and 2nd-graders' spatial learning. *Mind, Brain, and Education*, *7*(3), 196-199. https://doi.org/10.1111/mbe.12027

Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics anxiety in young children: An exploratory study. *The Journal of Experimental Education*, *81*(4), 538-555. https://doi.org/10.1080/00220973.2012.727888

Harth, H. (2018). *An analysis of the teaching of introductory statistics at university in "context"* [Doctoral dissertation, Loughborough University]. https://doi.org/10.26174/thesis.lboro.8246996.v1

He, X., Ren, M., & Ren, Y. (2025). The impact of mathematical anxiety and working memory on children's approximate number system and the n-back training effect. *Acta Psychologica*, *260*, Article 105330. https://doi.org/10.1016/j.actpsy.2025.105330

Henn, P., Gallagher, A. G., Nugent, E., Seymour, N. E., Haluck, R. S., Hseino, H., Traynor, O., & Neary, P. C. (2018). Visual spatial ability for surgical trainees: Implications for learning endoscopic, laparoscopic surgery and other image-guided procedures. *Surgical Endoscopy*, *32*(8), 3634-3639. https://doi.org/10.1007/s00464-018-6094-3

Hinze, S. R., Williamson, V. M., Shultz, M. J., Williamson, K. C., Deslongchamps, G., & Rapp, D. N. (2013). When do spatial abilities support student comprehension of stem visualizations? *Cognitive Processing*, *14*(2), 129-142. https://doi.org/10.1007/s103 39-013-0539-3

Hoegler, S., & Nelson, M. (2018). The influence of anxiety and self-efficacy on statistics performance: A path analysis. *Psi Chi Journal of Psychological Research*, *23*(5), 364-375. https://doi.org/10.24839/2325-7342.JN23.5.364

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment*, *10*(2), 178-182. https://doi.org/10.1177/1073191103010002008

Huang, F., Zheng, S., Fu, P., Tian, Q., Chen, Y., Jiang, Q., & Liao, M. (2023). Distinct classes of statistical anxiety: Latent profile and network psychometrics analysis of university students. *Psychology Research and Behavior Management*, *16*, 2787-2802. https://doi.org/10.2147/PRBM.S417887

Husereau, T., Cousineau, D., Zang, S., & Gibeau, R.-M. (2024). A compendium of common heuristics, misconceptions, and biased reasoning used in statistical thinking. *The Quantitative Methods for Psychology*, *20*(1), 57-75. https://doi.org/10.20982/tqmp.20.1.p057

Jaeger, A. J., Taylor, A. R., & Wiley, J. (2016). When, and for whom, analogies help: The role of spatial skills and interleaved presentation. *Journal of Educational Psychology*, *108*(8), 1121-1139. https://doi.org/10.1037/edu0000121

Jordan, J.-A., McGladdery, G., & Dyer, K. (2014). Dyslexia in higher education: Implications for maths anxiety, statistics anxiety and psychological well-being. *Dyslexia*, *20*(3), 225-240. https://doi.org/10.1002/dys.1478

Kearney-Ramos, T. E., Fausett, J. S., Gess, J. L., Reno, A., Peraza, J., Kilts, C. D., & James, G. A. (2014). Merging clinical neuropsychology and functional neuroimaging to evaluate the construct validity and neural network engagement of the n-Back task. *Journal of the International Neuropsychological Society*, *20*(7), 736-750. https://doi.org/10.1017/S135561771400054X

Kulikowski, K., & Potasz-Kulikowska, K. (2016). Can we measure working memory via the Internet? The reliability and factorial validity of an online n-back task. *Polish Psychological Bulletin, 47*(1), 51-61. https://doi.org/10.1515/ppb-2016-0006

Kyttälä, M., & Björn, P. M. (2014). The role of literacy skills in adolescents' mathematics word problem performance: Controlling for visuo-spatial ability and mathematics anxiety. *Learning and Individual Differences*, *29*, 59-66. https://doi.org/10.1016/j.lindif.2013.10.010

Laasonen, M., Lahti-Nuuttila, P., Leppämäki, S., Tani, P., Wikgren, J., Harno, H., Oksanen-Hennah, H., Pothos, E., Cleeremans, A., Dye, M. W. G., Cousineau, D., & Hokkanen, L. (2020). Project DyAdd: Non-linguistic theories of dyslexia predict intelligence. *Frontiers in Human Neuroscience*, *14*, 1-12. https://doi.org/10.3389/fnhum.2020.00316

Lalande, D., Cantinotti, M., Williot, A., Gagnon, J., & Cousineau, D. (2019). Three pathways from achievement goals to academic performance in an undergraduate statistics course. *Statistics Education Research Journal*, *18*(1), 94-105. https://doi.org/10.52041/serj.v18i1.152

Lauer, J. E., Esposito, A. G., & Bauer, P. J. (2018). Domain-specific anxiety relates to children's math and spatial performance. *Developmental Psychology*, *54*(11), 2126-2138.

https://doi.org/10.1037/dev0000605

Laurencelle, L., & Cousineau, D. (2022). Analysis of proportions using arcsine transform with any experimental design. *Frontiers in Psychology*, *13*, 1-12. https://doi.org/10.3389/fpsyg.2022.1045436

Lee, K.-M., & Kang, S.-Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, *83*(3), B63-B68. https://doi.org/10.1016/S0010-0277(02)00010-0

Lewin, S., Bosch-Capblanch, X., Oliver, S., Akl, E. A., Vist, G. E., Lavis, J. N., Ghersi, D., Røttingen, J.-A., Steinmann, P., Gulmezoglu, M., Tugwell, P., El-Jardali, F., & Haines, A. (2012). Guidance for evidence-informed policies about health systems: Assessing how much confidence to place in the research evidence. *PLOS Medicine*, *9*(3), 1-8. https://doi.org/10.1371/journal.pmed.1001187

Lisica, D., Koso-Drljević, M., Stürmer, B., Džubur, A., & Valt, C. (2022). Working memory impairment in relation to the severity of anxiety symptoms. *Cognition and Emotion*, *36*(6), 1093-1108. https://doi.org/10.1080/02699931.2022.2081535

Lowrie, T., & Logan, T. (2023). Spatial visualization supports students' math: Mechanisms for spatial transfer. *Journal of Intelligence*, *11*(6), Article 6. https://doi.org/10.3390/jintelligence11060127

Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial Anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition*, *4*(3), Article 3. https://doi.org/10.5964/jnc.v4i3.154

Macher, D., Paechter, M., Papousek, I., Ruggeri, K., Freudenthaler, H. H., & Arendasy, M.

(2013). Statistics anxiety, state anxiety during an examination, and academic achievement. *British Journal of Educational Psychology*, *83*(4), 535-549. https://doi.org/10.1111/j.2044-8279.2012.02081.x

Macher, D., Papousek, I., Ruggeri, K., & Paechter, M. (2015). Statistics anxiety and performance: Blessings in disguise. *Frontiers in Psychology*, *6*, Article 1116. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01116

Mackinnon, S. P., Alexander, S. M., Chen, R., Cribbie, R. A., Flett, G. L., & Hill, T. G. (2025). Perfectionism, anxiety sensitivity, and statistics anxiety: A test of the vulnerability-stress model using a 2-wave longitudinal study. *Journal of Research in Personality, 116*, Article 104611. https://doi.org/10.1016/j.jrp.2025.104611

Mandap, C. M. (2016). Examining gender differences in statistics anxiety among college students. *International Journal of Education and Research*, *4*(6), 357-366. https://www.ijern.com/journal/2016/June-2016/31.pdf

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17-24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

McIntee, S.-E., Goulet-Pelletier, J.-C., Williot, A., Deck-Léger, E., Lalande, D., Cantinotti, M., & Cousineau, D. (2022). (Mal)adaptive cognitions as predictors of statistics anxiety. *Statistics Education Research Journal*, *21*(1), Article 1. https://doi.org/10.52041/serj.v21i1.364

McMahon, P., Zhang, T., & Dwight, R. (2020). Requirements for big data adoption for railway asset management. *IEEE Access*, *8*, 15543-15564. https://doi.org/10.1109/ACCESS.2020.2967436

Miller, H. E., Kirkorian, H. L., & Simmering, V. R. (2020). Using eye-tracking to understand relations between visual attention and language in children's spatial skills. *Cognitive Psychology*, *117*, 1-36. https://doi.org/10.1016/j.cogpsych.2019.101264

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*(3), 272-277. https://doi.org/10.1093/oxfordjournals.bmb.a070866

Moreira Da Silva, M. P., & Pinto, S. S. (2014). Teaching Statistics Through Learning Projects. *Statistics Education Research Journal*, *13*(2), 177-186. https://doi.org/10.52041/serj.v13i2.289

Nesbit, R. J., & Bourne, V. J. (2018). Statistics Anxiety Rating Scale (STARS) use in Psychology students: A review and analysis with an undergraduate sample. *Psychology Teaching Review*, *24*(2), 101-110. https://eric.ed.gov/?id=EJ1196472

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The Single-Item Math Anxiety scale (SIMA): An alternative way of measuring mathematical anxiety. *Personality and Individual Differences*, *60*, S75-S76. https://doi.org/10.1016/j.paid.2013.07.339

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education*, *29*(1), 3-19. https://doi.org/10.1080/0260293042000160384

Onwuegbuzie, A. J., Da Ros, D., & Ryan, J. M. (1997). The components of statistics anxiety: A phenomenological study. *Focus on Learning Problems in Mathematics*, *19*(4), 11-35. https://eric.ed.gov/?id=EJ558838

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments--a comprehensive review of the literature. *Teaching in Higher Education*, *8*(2), 195-209. https://doi.org/10.1080/1356251032000052447

Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., & Papousek, I. (2017). Mathematics anxiety and statistics anxiety. Shared but also unshared components and antagonistic contributions to performance in statistics. *Frontiers in Psychology*, *8*, 1-13. https://doi.org/10.3389/fpsyg.2017.01196

Peng, P., & Ma, C. (2025). The role of domain-specific working memory and emotion regulation in the mathematics anxiety-performance relation among upper elementary students. *Developmental Psychology*. Advance online publication. https://doi.org/10.1037/dev0002081

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110-122. https://doi.org/10.1016/j.lindif.2009.10.005

Ramful, A., Lowrie, T., & Logan, T. (2017). Measurement of spatial ability: Construction and validation of the Spatial Reasoning Instrument for Middle School students. *Journal of Psychoeducational Assessment*, *35*(7), 709-727. https://doi.org/10.1177/0734282916659207

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology, 141*, 83-100. http://dx.doi.org/10.1016/j.jecp.2015.07.014

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2012). Spatial anxiety relates to spatial abilities as a function of working memory in children. *Quarterly Journal of Experimental Psychology*, *65*(3), 474-487. https://doi.org/10.1080/17470218.2011.616214

Rittle-Johnson, B., Zippert, E. L., & Boice, K. L. (2018). The roles of patterning and spatial skills in early mathematics development. *Early Childhood Research Quarterly*, *46*, 166-

178. https://doi.org/10.1016/j.ecresq.2018.03.006

Rodán, A., Montoro, P. R., Martínez-Molina, A., & Contreras, M. J. (2022). Effectiveness of spatial training in elementary and secondary school: Everyone learns. *Educación XX1*, *25*(1), 381-406. https://doi.org/10.5944/educXX1.30100

Rouet, J.-F., Vörös, Z., & Pléh, C. (2012). Incidental learning of links during navigation: The role of visuo-spatial capacity. *Behaviour & Information Technology*, *31*(1), 71-81. https://doi.org/10.1080/0144929X.2011.604103

Ruggeri, K., Dempster, M., & Hanna, D. (2011). The impact of misunderstanding the nature of statistics. *Psychology Teaching Review*, *17*(1), 35-40. https://eric.ed.gov/?id=EJ932187

Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 1-15. https://doi.org/10.1080/10705511.2019.1611434

Sokolowski, H. M., Hawes, Z., & Lyons, I. M. (2019). What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition*, *182*, 193-212. https://doi.org/10.1016/j.cognition.2018.10.005

Sorby, S., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes.' *Learning and Individual Differences*, *67*, 209-222. https://doi.org/10.1016/j.lindif.2018.09.001

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352-402. https://doi.org/10.1037/a0028446

Valle, N., Antonenko, P., Valle, D., Sommer, M., Huggins-Manley, A. C., Dawson, K., Kim, D., & Baiser, B. (2021). Predict or describe? How learning analytics dashboard design influences motivation and statistics anxiety in an online statistics course. *Educational Technology Research and Development*, *69*(3), 1405-1431. https://doi.org/10.1007/s11423-021-09998-z

van de Weijer-Bergsma, E., Van Luit, J. E. H., & Moeller, K. (2022). Sex differences in the association of math achievement with visual-spatial and verbal working memory: Does the type of math test matter? *British Journal of Psychology*, *113*(3), 798-819. https://doi.org/10.1111/bjop.12562

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicothema*, *20*(1), 174-180. https://reunido.uniovi.es/index.php/PST/article/view/8638

Villarejo-Ramos, Á. F., Cabrera-Sánchez, J.-P., Lara-Rubio, J., & Liébana-Cabanillas, F. (2021). Predicting Big Data adoption in companies with an explanatory and predictive model. *Frontiers in Psychology*, *12*, 1-12. https://doi.org/10.3389/fpsyg.2021.651398

Vytal, K., Cornwell, B., Arkin, N., Letkiewicz, A., & Grillon, C. (2013). The complex interaction between anxiety and cognition: Insight from spatial and verbal working memory. *Frontiers in Human Neuroscience*, *7*, 1-11. https://doi.org/10.3389/fnhum.2013.00093

Wang, S., Hu, B. Y., & Zhang, X. (2021). Kindergarteners' spatial skills and their reading and math achievement in second grade. *Early Childhood Research Quarterly*, *57*, 156-166. https://doi.org/10.1016/j.ecresq.2021.06.002

Williams, A. S. (2015). Statistics anxiety and worry: The roles of worry beliefs, negative problem orientation, and cognitive avoidance. *Statistics Education Research Journal*, *14*(2), 53-75. https://doi.org/10.52041/serj.v14i2.261

Yang, W., Liu, H., Chen, N., Xu, P., & Lin, X. (2020). Is early spatial skills training effective? A meta-analysis. *Frontiers in Psychology*, *11*, 1-15. https://doi.org/10.3389/fpsyg.2020.01938

Zang, S., & Cousineau, D. (2025). Anchors and ratios to quantify and explain y-axis distortion effects in graphs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *51*(9), 1430-1452. https://doi.org/10.1037/xlm0001454

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology*, *61*(3), 319-328. https://doi.org/10.1111/j.2044-8279.1991.tb00989.x

Chapter 4: Study C

Uncovering Key Statistical Concepts from an Investigation of Four Standardized Tests

Reference

Gibeau, R.-M., & Cousineau, D. (2025). Uncovering key statistical concepts from an investigation of four standardized tests. *Teaching Statistics, 47*(3), 219–234. https://doi.org/10.1111/test.12410

Contribution

Study conceptualisation, ethical request, data collection, data analysis and redaction/revisions of the manuscript.

Highlights

- Ecologically valid sample; the participants are psychology professors who taught one or many statistics courses.
- Four questionnaires of statistics' concept understanding are used.
- Eighteen statistical concepts were identified across participants' responses.
- Noteworthy, most participants do not agree with the items' correct answer given by the creators of the four questionnaires.

Disclaimer

Minor changes were made for the present thesis that are not in the published article to improve the text based on the comments of the thesis committee members. Also, citation is performed according to the *Teaching Statistics* requirement (AMA rather than APA format).

Abstract

There are few standardized tests measuring student performance with regards to statistics. Only four tests have been proposed for college or university students. The goal of the present study is to investigate these tests. University professors or instructors experienced in teaching statistics were asked to list the concepts they think are being assessed by each item of the tests. A total of 708 responses were obtained from 42 participants. Using thematic analysis, 18 fundamental statistical concepts were identified. Unplanned analyses on participants' scores were also computed. The results suggest that there is no consensus on some of the items' correct answers. The study has practical implications for teaching statistics, from learning goals to assessment methods.

Keywords: assessment, standardized statistics tests, statistics education

1 Introduction

In this Big Data Era, statistical information is everywhere.[41, 57] This information is used to promote certain ideologies, justify decisions, encourage or discourage people to adopt a certain behavior, or simply to share events or conditions happening in the world.[51, 52] As such, understanding the meaning and implications of this statistical information is a prerequisite for informed exchanges. However, it is commonly reported that statistics are among the most difficult subjects to teach and to learn.[14, 15, 26, 40] Over the last few decades, many researchers in education and psychology have investigated why the teaching and learning of statistics are so challenging and how they can be improved.[5, 19, 24, 44]

1.1 Research avenues to improve statistics education

To circumvent these difficulties, some research examined the teaching approach and the teacher's characteristics, and how they influence students' learning.[29, 34] Their work suggests that the use of humor by instructors facilitates students' understanding, which in turn may improve their evaluation of the class and of the instructor.[3, 20] Others argued that manipulating real-world data, doing demonstrations, and allowing students to manipulate data and probabilities help to develop statistical reasoning.[45, 54, 59]

Instead of alleviating the difficulties, other studies tried to identify the stumbling blocks which impede learning. Kahneman and Tversky[35, 55] and Konold[37] started this line of research by assuming that humans have heuristic modes of thinking (defined as simplistic ways of reasoning) and that these heuristics (reviewed in[32]) are keeping them from reasoning statistically, even after formal instruction. Kahneman and Tversky[35] examined people's ability to understand and manipulate probabilistic information. The authors concluded that heuristics are adopted when facing uncertainty. Furthermore, when researchers questioned students who correctly solved a problem, it was found that the correct reasoning did not underlie their correct response.[38]

A third avenue of research, one that has been seldom taken, is to identify the core concepts underlying statistical understanding and which representations would best make them accessible to learners. Rather than alleviating learning difficulties or circumventing inadequate reasoning, this line of research takes the complementary approach of identifying what has to be conveyed to learners (i.e., the learning goals). Gigerenzer,[28] in particular, argued that probabilities (expressed with percentages) do not fit the learners' internal representations, which may be more aptly evoked with frequencies (expressed with counts). Consequently, identifying a list of the core concepts and how they can be distinguished from heuristic thinking would be a useful step in developing valid and discriminant assessment tools that go beyond mere memorization capacity and procedural skills. Indeed, it may be beneficial to invest efforts in developing assessments of conceptual understanding of statistics.

1.2 Procedural skills and conceptual understanding

The term *procedural skills* can be defined as "the ability to execute action sequences to solve problems"[47] (p. 346). In statistics, an example of procedural skills is the ability to calculate the mean given a set of values by applying a formula. To evaluate this type of learning, it is suggested to use typical problems for which students have step-by-step procedures to solve them. With technological advancements, specifically the existence of statistics software, it is argued that it is no longer necessary and even useful to teach students how to compute certain statistics by hand.[9, 12, 43] It is instead recommended to teach students how to use technology, how to choose the appropriate tool, how to interpret the outputs and how to communicate the results. The motivation to move from learning internalized procedural skills to learning technologically supported procedural skills, according to some authors, is to reduce cognitive effort to allow the learners to invest more cognitive capacity onto conceptual learning.

A *concept* can be defined as how the understanding of something is internalized (typically a word or a symbol). It generally includes the characteristics of that thing as well as how it associates with other things.[36] As such, *conceptual understanding* can be defined as the understanding of the representation of similarities and differences (or characteristics) of elements and the association between them or how they influence each other.[47] For example, the concept *mean* is the representation of a response that is representative of a group of individual responses that are both similar and different from each other; a change in their individual values will change the value of the representative response and the visual representation of it. To evaluate this type of learning, it is suggested to use new and original problems for which students do not have a step-by-step procedure to solve them.[47] To complete these assessments, students must rely on conceptual knowledge to create an appropriate procedure that will solve the problem.

1.3 Evaluating students in statistics courses

According to Allen,[2] instructors too often assess students' memorization capacity based on how well they can remember formulas, rules, definitions or assumptions. However, focusing on students' conceptual understanding would influence their learning strategies and foster transferable knowledge.[11, 33] For this reason, some researchers argue that instructors do not assess students correctly.[12, 22] It can even be argued that the assessments used in introductory statistics courses are not valid for evaluating conceptual understanding. For example, anyone seeing the response choice “Correlation is not causation” will select it simply because it has been repeated frequently and memorized.

On a related note, statistics anxiety questionnaires indicate that the anxiety of interpreting statistical results is a major component of the difficulty to understand statistics (an example item is “how anxious are you when interpreting the meaning of a table”).[10, 27, 56] Assessing students in an appropriate way, avoiding rote learning and computation, could reduce anxiety which, according to some, is the most detrimental influencer of their performance.[5, 6, 10, 14, 26, 31, 44]

To assess students' understanding of statistics, it has been proposed to use specific items that would measure three basic learning goals: statistical literacy, statistical reasoning, and statistical thinking.[19, 30] The definition of statistical literacy varies in the literature. Aspects that often come up are the ability to read statistical information, critically evaluate it, form a reflection about such information and communicate this reflection.[50] To evaluate students' statistical literacy, instructors can use the terms *identify*, *describe*, *rephrase*, *translate*, *interpret*, and *read*.[19] Statistical reasoning is defined as the way people reason with statistical information, which involves interpreting results and graphs, and understanding the tools available.[23] To assess statistical reasoning, the terms *why*, *how* or *explain the process* can be used.[19] Finally, statistical thinking is defined as “what a statistician does”[13] (p. 4). In other words, it is the ability to explore data, to go beyond the teachings, being naturally curious about statistical problems and to see the process of doing statistics (analyses, interpretations, etc.) as a whole. The terms *apply*, *criticize*, *evaluate*, and *generalize* are thought to measure this learning goal.[19] It can be argued that the first and second learning goals are a way to develop students' conceptual understanding, while statistical thinking is a way to develop students' procedural skills.

Other research, based on Garfield's[24] assessment framework, provided ways to evaluate students depending on the instructor's purpose for the assessment.[12, 26] Based on this work, instructors can choose between eight different ways to evaluate their students: homework, quizzes and exams, projects, critiques and writing assessments, lab reports, collaborative assessment tasks, minute papers, and student reflections. Additionally, instructors have three teaching strategies at their disposal according to Marson,[39] namely, immediate feedback, repetition, and original data (collected by learners). The author reported that immediate feedback helps students the most in their learning of statistics. This beneficial effect is presumably caused by students' awareness of their use of invalid heuristics. As interesting as these evaluation methods are, they do not concretely solve the problem of evaluating students' conceptual understanding instead of procedural skills.

1.4 Assessments in statistics education

As mentioned above, conceptual understanding includes statistical literacy and statistical reasoning. To our knowledge, there are four tests that claim to measure one of these two components, described in the Methods section in more detail: the *Statistical Reasoning Assessment* (SRA),[25] the *Statistics Concept Inventory* (SCI),[2] the *Comprehensive Assessment of Outcomes for a First Course in Statistics* (CAOS),[18] and the *Statistical Reasoning in Biology Concept Inventory* (SRBCI).[17] However, when examining the tests' items, the underlying concept each item is measuring is unspecified and often unclear; the developers do not clearly state what each item measures. They provide general topics or dimensions under which items were created without specifying the concepts measured. Some items, across all four tests, seem to measure students' rote learning and computational skills (procedural skills) rather than conceptual learning. Thus, they may not be valid instruments for measuring students' conceptual understanding of statistics.

There are other instruments that have been developed to measure statistics learning.[49, 58] Some were developed for children[58] or include questions on research methods.[49] They often confuse performance with learning—because a student with a good performance does not necessarily understand the subject—and deviate very little from problems given in statistics classes. For example, Sabbag and colleagues[49] are focused on measuring statistical literacy and statistical reasoning and the relation between these two learning goals; it is therefore mute as to the concepts underpinning these goals.

The aim of the present study is to investigate the following question: which concepts do statistics instructors think are being measured in standardized statistical instruments? An additional exploratory research question is: Do statistics instructors agree on these concepts and on the items' correct answer? In other words, the goal is to identify the underlying statistical concepts measured in these four tests through qualitative and quantitative data using statistics instructors because they are the ones teaching the subject and administering those tests. Their perspective is essential for understanding the tests mentioned above. Indeed, not being the developers of the tests and having pedagogical knowledge and experience in statistics teaching, make them the sample of choice. They understand how to construct evaluations based on their teaching and, inversely, what has been taught based on the evaluations. Their answers will clarify the content of those tests. As such, a sample of experienced statistics educators was invited to list which concepts are being tested in each item (qualitative) and optionally to answer the tests' items (quantitative). With their qualitative answers, the goal is to identify the building blocks of the statistical problems.

2 Methods

Prior to data collection, this study underwent ethical review and was approved by the *Ethics Research Board* of the University of Ottawa.

2.1 Participants

In non-math university programs, like psychology, it is the department's professors that are teaching the statistics courses and not statisticians or mathematicians. To increase the study's ecological validity, professors or instructors in psychology who have taught one or more statistics course(s) prior to the study were recruited by an invitation email sent through the mailing lists of the *Canadian Society for Brain, Behavior and Cognitive Science* (CSBBCS) and the *Société Québécoise pour la Recherche en Psychologie* (SQRP). In addition, all psychology professors at the authors' university and their past collaborators were contacted. Some reported in the comments that they have taught statistics classes many times; one also mentioned that he had just retired. In this invitation email, details about the aim of the study, what is required from participants, and the URL link to the online questionnaire were provided.

The chosen population is better suited than any other population to participate in this study for two main reasons. Firstly, they are psychology professors, implying that they are frequent users of statistics for their own research projects. They likely appreciate the usefulness of such tools in research, the omnipresence of such information in daily life, and the necessity of understanding statistics and forming statistically qualified citizens. Secondly, they have taught one or more statistics course(s) to university students. This means that they have some pedagogical comprehension of the subject, they have experience with their field's curriculum and learning goals, and they are aware of concerns, fears and capabilities of students. These criteria ensure that participants are not statisticians and ensure that they are not influenced by the hindsight bias (the tendency of considering an information as being predictable and obvious after it was known[1]), in which case it would decrease the study's validity by focusing on experts instead of instructors.

No target sample size was a priori decided. The goal was simply to have as many participants as possible for the duration of one semester considering that this is a specific and limited population. In total, 42 participants completed the study.

2.2 Material

The four standardized tests listed above were assembled into one questionnaire in Qualtrics (a software used to create and distribute surveys). They are all multiple-choice tests measuring statistical understanding. In total, there were 97 items. The tests are detailed here.

2.2.1 SRA

The *Statistics Reasoning Assessment* is composed of 20 multiple-choice items (available in the validation article[25]). Most items are short scenarios with one or more correct answers. As an example, the second item is:

"The following message is printed on a bottle of prescription medication:

WARNING: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult your physician.

Which of the following is the best interpretation of this warning?"

The response choices are: "don't use the medication on your skin, there's a good chance of developing a rash," "for application to the skin, apply only 15% of the recommended dose," "if a rash develops, it will probably involve only 15% of the skin," "about 15 of 100 people who use this medication develop a rash," and "there is hardly a chance of getting a rash using this medication." The correct answer is the fourth response choice.

The instrument was developed by Konold[37] for high school students and validated by Garfield.[25] It was developed in three steps: (i) experts reviewed the items for content validity, (ii) the items were administered to a group of students who answered open-ended questions on the justification of their responses, which were used to form the SRA's response choices, and (iii) multiple pilot tests were run.

The SRA aims to measure two dimensions, *Correct Reasoning Skills* and *Misconceptions*, which were divided into eight skills according to the authors. The first dimension includes: Correct interpretation of probabilities, understanding how to select the right average, correct computation of probabilities (further includes understanding probabilities as ratios and using combinatorial reasoning), understanding independence, understanding sampling variability, distinguishing between correlation and causation, correctly interpreting two-way tables, and understanding the importance of large samples. The second dimension includes: Misconceptions involving averages (further includes the belief that averages are the most common number, the failure to consider outliers when computing the mean, group comparison using their averages, and confusion between mean and median), outcome orientation misconception, misconception about good samples needing a high percentage of the population, law of small numbers, representativeness misconception, correlation implies causation, equiprobability bias, and groups can only be compared if they have the same sample size. Notably, each skill is not measured by one or multiple question(s), but by one or multiple response choice(s). For example, the correct interpretation of probabilities is measured by response choice *d* of item 2 and item 3.

The internal consistency of the SRA has raised some issues and “yielded incomplete results”[25] (p. 31) but good test–retest reliability was observed for participants' correct responses score ($r = .70$) and incorrect responses score ($r = .75$). Criterion-related validity was also examined by asking students enrolled in an introductory statistics course to complete the test. Students' scores on the SRA were then correlated with their grade on assignments. The author of the validation article mentioned that the correlations were low, without providing the coefficients.

2.2.2 SCI

The *Statistics Concept Inventory* (SCI) has 25 multiple-choice items. They are short scenarios with at least one correct answer. The SCI was developed as part of a doctoral dissertation[2] (the items are also available in the dissertation). Most participants were engineering students, some were psychology students, and a small number of participants were in other programs. As an example, the first item is:

> "A certain diet plan claims that subjects lose an average of 20 pounds in 6 months on their plan. A dietician wishes to test this claim and recruits 15 people to participate in an experiment. Their weight is measured before and after the 6-month period. Which is the appropriate test statistic to test the diet company's claim?"

The response choices are: "two-sample *Z* test," "paired comparison *t* test," and "two-sample *t* test." The correct answer is the second response choice.

The items and the response choices were created based on textbooks, journal articles, and personal experiences. Also, the items are divided into four topics: inferential statistics, probability, descriptive statistics, and graphs. Through a process of focus groups, analysis of correct and incorrect answers, and expert opinions, the author revised the items of the questionnaire. The SCI was tested and revised six times from Fall 2002 to Spring 2006 with engineering and mathematics students. To keep the present text as short as possible, only the initial and final versions are discussed below.

The author reported the alpha and omega reliability indices with the initial 38-item SCI (Cronbach's $\alpha = .77$ and McDonald's $\Omega = .89$[4, 16]). An exploratory factor analysis suggested three solutions: a five-factor model with an oblique rotation where the loadings vary between .30 and .66 with one cross-loading; a five-factor model with an orthogonal rotation where the loadings vary between .30 and .64 with four cross-loadings; and a four-factor model without rotation where the loadings vary between .30 and .59 with one cross-loading. A total of 13 items were removed because previous research suggested a smaller number of factors compared to the first version of the SCI, which questioned face validity. After removing these 13 items, resulting in a 25-item test, the author did a second confirmatory factor analysis which supported a one-factor model (GFI = .92, PGFI = .84, RMSEA = .03). Next, the reliability of the 25-item SCI was examined ($\alpha = .76$). Finally, content validity was assessed through student interviews and faculty surveys. The author then claimed that this modified SCI would measure 14 important statistical concepts.

2.2.3 CAOS

The *Comprehensive Assessment of Outcomes for a First Course in Statistics* (CAOS) has 40 multiple-choice items. They are short scenarios with at least one correct answer. This instrument was developed as part of the *Assessment Resource Tools for Improving Statistical Thinking* (ARTIST) project, which aimed at providing items and instruments for statistics education.[18] It was developed for post-secondary students enrolled in an introductory statistics course. The items are available upon request to the first author of.[18] As an example, the seventh item is:

> "A recent research study randomly divided participants into groups who were given different levels of Vitamin E to take daily. One group received only a placebo pill. The research study followed the participants for eight years to see how many developed a particular type of cancer during that period. Which of the following responses gives the best explanation as to the purpose of randomization in this study?"

The response choices are: "to increase the accuracy of the research results," "to ensure that all potential cancer patients had an equal chance of being selected for the study," "to reduce the amount of sampling error," "to produce treatment groups with similar characteristics," and "to prevent skewness in the results." The correct answer is the fourth response choice.

The CAOS was developed over the course of three years. First, it was decided to focus the test on variability through 10 topics. Then, the authors assembled items from statistics instructors and created additional items for the topics that were not covered. The initial version had 34 items. This version was pilot tested with students enrolled in an introductory statistics course. The data allowed the authors to revise this version, which resulted in a 37-item test. The CAOS was revised and tested two more times. The final version has 40 items divided into 10 topics: data collection and design, descriptive statistics, graphs, boxplots, normal distribution, bivariate data, probability, sampling variability, confidence intervals, and significance testing.

Reliability of this final version was examined using Cronbach's alpha. The results suggest a good internal consistency with college students. Content validity was examined by asking 18 members of the *Consortium for the Advancement of Undergraduate Statistics Education* (CAUSE), who are both statisticians and statistics instructors, to act as experts and rate the items. These members revised the test, answered a few questions about its validity and unanimously agreed with the statements "CAOS measures important outcomes that are common to most first courses in statistics" and "CAOS measures outcomes for which I would be disappointed if they were not achieved by students who succeed in my statistics courses" (p. 31). Also, 94% of these experts agreed with "CAOS measures important outcomes that are common to most first courses in statistics"[18] (p. 31).

2.2.4 SRBCI

The *Statistical Reasoning in Biology Concept Inventory* (SRBCI) is a 12-item multiple-choice measure. The items are divided into three blocks of four items each related to the same graphic and scenario. This instrument was developed for assessing undergraduate biology students' understanding of statistics.[17] The items are available on this website: https://q4b.biology.ubc.ca/concept-inventories/statistical-reasoning-in-biology/. As an example, the ninth item is:

> "You tested whether temperature affected the growth of sunflowers over a 6-month period after germination. You tested 24 plants in each of three treatments (12°C, 13°C, 14°C). You recorded the height (cm) of each plant 6 months after germination, and then calculated averages (means) and a measure of variation around the averages (95% confidence intervals).
>
> At 12°C, plants grew 114 cm ± 3 cm, at 13°C they grew 122 cm ± 4 cm, and at 14°C they grew 120 cm ± 8 cm. Which of the following is the most accurate interpretation of these results?"

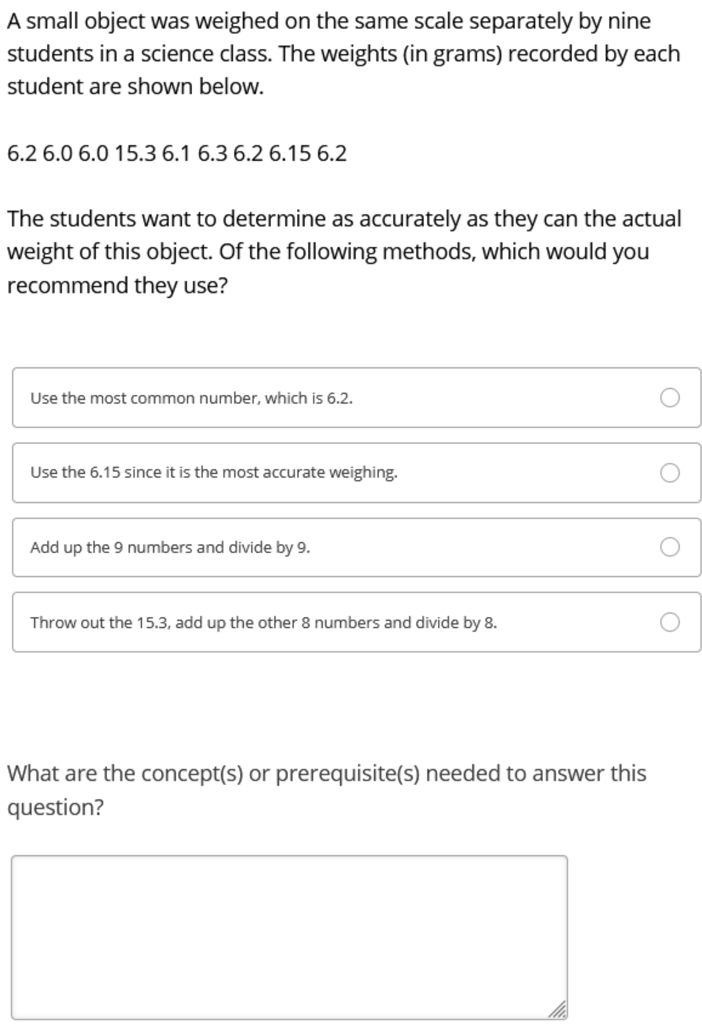
The response choices are: "Temperature did affect growth rate because plants grew significantly taller at 13°C than at 12°C," "temperature did not affect growth rate because plants grew tallest in the middle temperature (13°C)", "temperature did affect growth rate because the average (mean) growth was different in each of the three temperatures," and "temperature did not affect growth rate because plants grown in the coolest (12°C) and hottest (14°C) temperatures did not grow significantly differently." The correct answer is the first choice.

The items are divided into four "core conceptual grouping[s]"[17] (p. 4): repeatability of results, variation in data, hypotheses and predictions, and sample size. To create the items, the authors first examined the literature, looked at statistics exams and assignments and surveyed faculty members on important statistical concepts. The results indicate that there are four key concepts, which are the focus of the SRBCI: data variability, replication of results, hypotheses and predictions, and sample size. Then, the authors created three items for each key concept for a total of 12. The first and only version was tested with science students using interviews. Validity of this version was tested with expert revision. No clear indication of reliability or validity is provided. However, the authors ran a Rasch model analysis[7, 46, 48] where the results suggest that the data fits the one-dimensional model well. Pairwise item correlations were calculated and were all within recommendations except for the correlations between item 1 and item 6, between item 3 and item 11, and between item 7 and item 8. Then, the infit and outfit mean-square's measures of fit of each item were calculated. These varied between 0.6 and 1.4, which is within recommendations. Next, the Pearson product–moment correlations were computed between participants' score on the SRBCI and the Rasch model estimates. The authors reported almost linear correlations. Lastly, item difficulty was examined. It was concluded that the SRBCI is an appropriate instrument for evaluating students' statistical reasoning.

2.3 Procedure

The four tests were assembled into a single survey. The SRA's 20 items were divided into 18 blocks of item(s) as items 5 and 6 share the same scenario, and similarly for items 9 and 10. In the SCI, 25 blocks of one item each were formed since each item has its own scenario. In the CAOS, the 40 items were divided into 25 blocks of item(s) with the following items grouped together because they have the same scenario: 1 and 2, 3 to 5, 8 to 10, 11 to 13, 14 and 15, 23 and 24, 25 to 27, 28 to 31, and finally 34 and 35. Concerning the SRBCI, as said above, its 12 items are grouped into 3 blocks of 4 items each: items 1 to 4, 5 to 8, and 9 to 12. Consequently, all items were spread over 71 blocks.

Because 71 is a high number of blocks, a random subset of 20 of these blocks were presented to each participant. All 71 blocks were divided randomly across participants. Further, the order of the selected blocks was random. Participants were uninformed to which items of which scale they answered. However, participants were informed that they did not have to answer the items. What really mattered in each block was an open-ended question where they were asked, "What are the concept(s) or prerequisite(s) needed to answer this question?" Their answer to this question provided the qualitative data kept for the thematic analysis. A screen capture of a block is presented in Figure 1. After participants completed the questionnaire, their quantitative scores for each item were given to them along with the correct answer, based on the tests' developers. Participants received one point for each correctly answered item. Finally, they could provide comments on the last page.



A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.

6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2

The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?

- Use the most common number, which is 6.2.
- Use the 6.15 since it is the most accurate weighing.
- Add up the 9 numbers and divide by 9.
- Throw out the 15.3, add up the other 8 numbers and divide by 8.

What are the concept(s) or prerequisite(s) needed to answer this question?

FIGURE 1

The first item of the SRA with the added question.

2.4 Planned Analyses

2.4.1 Qualitative

The text entries of each block of item(s) are the focus of this research. To analyze them, a thematic analysis divided into 6 phases was done, as described by Braun and Clarke.[8] The first and last phases include familiarization with the dataset and reporting the themes that were identified. The second phase involves generating codes for each block of item(s). Those codes are the statistical concepts listed by participants. Some participants wrote complete sentences explaining their reasoning, while others only provided the concepts they thought were necessary to respond to the item. As such, the codes helped to differentiate between words that are pertinent to the research objectives and words that are not. The third phase concerns the identification of potential themes in each block of item(s). The themes are codes or groups of codes that could be united. In other words, the themes can be identified through patterns of responses, that is, statistical concepts that are often explicitly or implicitly mentioned by participants. In the fourth phase, the list of potential themes is revised. It is required that themes do not only reflect codes but are also pertinent to the objectives. Themes must reveal something meaningful about the data; they must reflect the entire dataset in a coherent way. Finally, in the fifth phase, the themes are defined and named so that they include all the codes. Note that the analysis was realized by the first author and revised by the second author. Also, this study is exploratory because there is no a priori framework from which one can rely on.

2.4.2 Quantitative

Participants' scores were calculated. Basic frequencies, sums, and means were used.

3 Results

3.1 Data Screening

Before data cleaning, 117 instructors clicked on the link. However, most of them did not finish the questionnaire. Specifically, some did not answer all open-ended questions and others simply did not answer any question. Only 42 participants, or 35.9%, who responded to all the 20 blocks were kept for analysis. Between six and 15 qualitative responses were recorded per item. Neither demographic data nor their number of years of experience as a statistics instructor was collected. Those 42 participants provided qualitative responses to 710 blocks with 130 missing responses on the tests' items: 169 in the SRA (response rate of 88.5%), 259 in the SCI (response rate of 83.0%), 247 in the CAOS (response rate of 84.3%) and 35 in the SRBCI (response rate of 79.5%); remember that the participants were told that responding to the items was optional. Tables S1 to S4 (available on OSF; https://osf.io/t6w7n/) show the frequencies of qualitative responses per block of item(s).

Three examples of responses are as follows. In response to the first item of the SRA: "In this question one would need to know how to identify a genuine outlier and further recognize that the mean is likely the most accurate answer. On a related note, we do not teach writing up statistics in stats class usually. This would be a great one to follow with asking them how they would write that up in a results section." For the items 34 and 35 of the CAOS, one response was: "understanding of random sampling and sampling distributions, the central limit theorem." Regarding the items 1 to 4 of the SRBCI, a participant wrote, "confidence intervals, sampling distributions, ANOVA, interpreting null effects."

Participants took on average 3 h and 53 min to complete the questionnaire. Since Qualtrics saves participants' progression, some must have completed it in an "on and off" manner or over multiple sessions, resulting in four durations exceeding 20 h. The smallest amount of time taken is 21 min and the highest amount of time taken is 43 h and 57 min. The median completion time is 55 min.

3.2 Themes Per Tests

3.2.1 Identification of codes

In the SRA, a total of 60 codes were identified for an average of three codes per block. In the SCI, 69 codes were identified with approximately three per item. For the CAOS, there are 87 codes in total, likewise for an average of three codes per block. Finally, in the SRBCI, 23 codes were identified in total for an average of eight codes per block. The tables of codes for each test

(Tables S5–S8) are available on OSF at https://osf.io/t6w7n/.

The codes of the SRA are somewhat similar to the 16 skills mentioned in the Methods section. Some other codes are very different from what is measured according to the authors. For example, item 24 of the CAOS is thought to measure understanding of statistical significance, statistical power, causality, type I and II errors. However, the authors mentioned that this item measure "understanding that an experimental design with random assignment supports causal inference"[18] (p. 44). Another example concerns items 5 to 8 of the SRBCI. Table S8 indicates that these items measure understanding of CIs, variance and variability, sample size, NHST, central limit theorem, and standard error. However, the authors mentioned that they measure understanding of "hypotheses and predictions" (items 5, 7, and 8), and "repeatability of results" (item 6)[17] (p. 4). These differences suggest that when one examines the items alone, the statistical concepts measured by those tests are not clear. The codes in bold represent procedural skills. As can be seen, some tests measure more procedural skills than others.

3.2.2 Potential Themes

The potential themes of each questionnaire across items are shown in Table **1**. The recurrence was calculated at the questionnaire level, meaning that 44% of the SRA's responses on the open-ended questions concern probability and trial independence. The SRA has eight themes, with three that have a recurrence of 14% and higher (i.e., *Probability and Trial Independence*, *Sample and Sampling*, and *Central Tendency*). The SCI has 12 themes, with three that have a recurrence above 10%. These are: *Central Tendency*, *Standard deviation* (SD)/*Variance*, and *Distribution* (*normal*, *sampling*, etc.). The CAOS (which is also the longest test with 40 items) has 13 different themes, with four that have a recurrence of 10% or more, namely, *Null-hypothesis Statistical Testing* (NHST)/*p-value*, *Distribution* (*normal*, *sampling*, etc.), *Correlation/Causality*, and *Error* (*sampling, types, standard*). Lastly, the SRBCI (the shortest test with 12 items) has only three themes (i.e., *NHST* with a recurrence of 33%, *Sample Size* with a recurrence of 28%, and *Confidence intervals* [*CIs*] with a recurrence of 28%).

TABLE 1. Potential themes of each questionnaire with their percentage of recurrence (percentages may add to more than 100% as some question items were identified with more than one theme)

SRA	SCI	CAOS	SRBCI
Probability & trial independence (44.1%)	Central tendency (14.9%)	**NHST/p-value (17.3%)**	**NHST (33.3%)**
Sample & Sampling (17.3%)	**SD/variance (14.9%)**	Distribution (normal, sampling, etc.; 17.0%)	Sample size (28.2%)
Central tendency (14.5%)	Distribution (normal, sampling, etc.; 14.2%)	**Correlation/causality (12.2%)**	**CIs (28.2%)**
Outlier (7.8%)	**Correlation (7.6%)**	Error (sampling, types, standard) (10.3%)	
Error & Bias (7.3%)	Probability (7.3%)	Graphs/visualization (9.6%)	
Sample size (6.2%)	Sample size (5.9%)	**Research methods/design (7.8%)**	
Statistical tests (6.2%)	**CIs (5.2%)**	Variation/variability (7.4%)	
Math (5.6%)	**NHST/p-value (5.2%)**	Central tendency (5.5%)	
	Outliers (4.5%)	Asymmetry (4.4%)	
	Rank order (4.5%)	Outliers (4.4%)	
	Graphs (4.2%)	**Descriptive statistics (4.1%)**	
	Error/bias (3.5%)	Probability (3.7%)	
		Regression (3.7%)	

Note. SD stands for standard deviation; NHST stands for Null Hypothesis Statistical Tests; CIs stands for Confidence Intervals. The procedural skills are in bold

Given the number of items, the SCI seems to be the one exploring the largest number of statistical concepts (12 themes for 25 items) whereas the SRBCI is the narrowest test (3 themes for 12 items). Two of the three themes in the SRBCI also happen to be high-level themes which need ample prior training to be mastered or understood (specifically *Confidence Intervals* and *Null Hypothesis Statistical Testing*). Also, the test least focused on procedural skills and most focused on conceptual understanding seems to be the SRA.

3.3 Themes Across Tests

The authors identified 994 codes from the participants' responses across all tests. At first, to construct this final list, only the tests' themes with a recurrence of 10% and higher were kept. Some of those themes were merged to form a unique theme. Ten different themes were identified. After discussion, it was realized that the themes with lower recurrence are not necessarily less important in statistics. Their low recurrences may have been caused by how the items were formulated and how the tests were developed. As such, after merging some themes together, the final list has 20 themes (including one theme not further delineated into multiple codes).

TABLE 2. Global themes across tests with percentage of recurrence

Themes	Recurrence
Probability & Trial independence	114 (11.6%)
Central tendency	98 (10.0%)
Distributions	97 (9.9%)
NHST	75 (7.6%)
Sample & Sampling	70 (7.1%)
Correlation & Causality	69 (7.0%)
Statistical tests & Computation	66 (6.7%)
Error & Biases	61 (6.2%)
Confidence intervals	48 (4.9%)
Descriptive statistics	48 (4.9%)
Graphs &Visualization	44 (4.5%)
Outliers	42 (4.3%)
Variability & Dispersion	35 (3.6%)
Research methods/designs	29 (3.0%)
Randomness & Randomization	25 (2.5%)
Asymmetry	18 (1.8%)
Representativity & Generalizability	12 (1.2%)
Central limit theorem	10 (1.0%)
Statistical power	8 (0.8%)
Unclassified	14 (1.4%)
Total	983 (100.0%)

Notes. The themes that were not further delineated into multiple codes include: (in)dependence/ codependence, linear relation, prediction, law of large numbers, interpretation, calibration, outcome. The procedural skills are in bold.

As can be seen, the tests have a few common themes: probability, central tendency, distributions, NHST, and samples. These themes are somewhat similar to the first dimension of the SRA and to the four dimensions of the SRBCI but are quite different from the topics covered in the SCI—which are mostly computational. Regarding the CAOS, some themes are somewhat similar, and others are quite different.

These 20 themes were then revised (see Table **3**). Through this revision, some themes were paired, some stayed the same, and others were renamed to ensure that they truly captured the data and the codes. In total, 18 revised themes were defined. These themes represent the concepts that are measured in the four tests (SRA, SCI, CAOS and SRBCI).

TABLE 3. Revised global themes across tests and their definition

Themes	Definitions
Probability & Randomness	Some events are deterministic, others random. Random events are unpredictable in detail, but some events are less random than others, in the sense that they are more likely than others. Random events can occur with varying degrees of certainty (quantification level of the concept). The degree of certainty of occurrence of a random event (Bayesian definition); establishing the frequency of its occurrence in identical reproductions (frequentist definition).
Central tendency	The population may have a more typical observation (conceptual level). It is possible to estimate the most frequent, the most central, or the least erroneous magnitude of random events (quantification level). These three estimates are slightly different and correspond to the mode, median and mean respectively. Populations also have theoretical central tendencies; we can only estimate these from a sample (population parameter level).
Distributions	Conceiving the repartition of distinct events (sampling distribution), the repartition of levels of inaccuracy (error distribution). Understanding that these random draws result in regularities (sampling laws). Conceiving the bell curve (normal distribution) as an example of a distribution.
NHST & p-value	NHST is a particular form of statistical inference (conceptual understanding). We can pose an a priori hypothesis about the result of a statistic, and state its "null" version. We choose a level of risk to adopt for the decision. There are several procedures (tests) for testing a null hypothesis, and these procedures are based on assumptions (quantification level). We can apply the procedure to obtain inferential statistics; establish the level of plausibility of the data when the null hypothesis is assumed (p-value); make a decision based on this level of plausibility.
Sample & Sampling	Randomly draw instances from the target population. For each draw, these instances will be different. There are many ways one can randomly draw instances from a population
Shared variance & Correlation	Two characteristics of a subject can be associated; these associations can be positive or negative (conceptual understanding). These associations are expressed as co-variation. These statistics are part of descriptive statistics but require bi-variate data. The degree of association can be quantified by covariation, and by correlation (level of quantification). These degrees of association can be compared. Bivariate populations also have a covariation/correlation (level of a population parameter).
Error & Its margins	Whenever a descriptive statistic is quantified, there ought to be some error in this quantification (conceptual understanding). This is due to chance and variability in the population examined, so the sample is more or less perfectly representative of the population. This sampling error leads to errors as soon as statistical results are produced from this sample. It may be in descriptive statistics (quantification level: estimation error, which can be quantified with the standard error and the precision interval) or in the hypothetical value given to a parameter (quantification level: inference error, which can be quantified with the confidence interval). It can also be described by a range of values adapted to a level of risk (quantification level: margin of error can be more or less narrow, depending on whether you want a margin of error that most probably contains the true value (high confidence) or fairly probably contains the true

	value (lower confidence). Any result is probably wrong to some extent, and so an estimate of the error should always accompany the result. These estimates can be quantified (standard error, confidence interval, or other technique; quantification level). These margins of error can be interpreted. Margins of error depend on the experimental design.
Descriptive statistics	A sample can be described. Central tendencies, variabilities & dispersions and asymmetries are possible descriptions, but there are others (e.g., quantiles and percentile ranks; conceptual understanding). These descriptive statistics are distinct from population parameters but often serve to approximately quantify these parameters when they are unknown (with a range of values). These descriptive statistics can be quantified and compared (quantification level). Populations also have other aspects (level of a population parameter).
Graphs &Visualization	Determining what is displayed on the graph; in particular, distinguishing between Venne diagrams, scatter graphs and summary graphs. Understanding the notion of dependent variable (the one to be read) and independent variable(s) (the one(s) that form(s) the data grouping(s)) or the notion of bivariate. Understanding whether what is displayed is raw data or descriptive statistics. When descriptive statistics are used, (i) determining whether margins of error accompany the results and (ii) being able to compare different statistics (conceptual understanding).
Outliers & Contamination	Sometimes, the sample is contaminated by elements (participants) who are not actually members of the target population. These contaminants may be visible (outlying) or undetectable (inlying). An absence or a low level of contamination is preferable (conceptual understanding).
Variability & Dispersion	Observations (participants) differ from one another, and the degree of difference may vary according to the population studied and the sample drawn from the population (conceptual understanding). This degree of difference is quantified by the typical deviation between observations. These typical deviations can be compared (quantification level). Populations also have theoretical dispersion trends; we can only estimate them from a sample (population parameter level).
Research, Research methods & Research designs	The conditions under which the data are obtained (e.g. whether subjects are measured once or several times; conceptual understanding). Data can be independent groups, repeated measures or paired. The specification is important to ensure that the data are valid (free of confounds), representative of the population and generalizable to that population (quantification level).
Asymmetry	Sampled scores can be asymmetrical, i.e., more frequent on one side of the distribution than the other, more long-range or more squashed. This asymmetry may be caused by a floor effect, a ceiling effect, contaminants and outliers, for example (conceptual understanding). This asymmetry can be quantified using the relevant formula (skewness or kurtosis); these figures can be compared (quantification level). Populations also have a theoretical asymmetry (level of a population parameter). A normal population has no asymmetry (its value is zero).
Population	Each study has one or more target groups of people, animals or objects. It is this (these) group(s) that is (are) typically recruited to participate. The results of a study are relevant to this (these) group(s) (conceptual understanding).
Modeling	The population and its relations are idealized by a modeling process. This process posits theoretical distributions, means, and/or correlations or covariations (conceptual understanding). Models built on data can be used to summarize the relations that exist between

	variables, and how they influence each other (quantification level).
Data & Observation	Data are observations made on subjects participating in the study. Data can be univariate (composed of a single number quantifying a single characteristic of the subject, such as empathy level), bivariate (composed of two numbers quantifying two characteristics of the subject, such as height and weight) or multivariate (two or more numbers quantifying two or more characteristics of the subject). Each individual piece of data is unpredictable, as the subject is chosen at random (conceptual understanding).
Decisions, Risks & Consequences	An informed decision must be based on considerations of risk levels and the consequences of these risks. For example, a decision to undergo ovarian ablation following a single positive diagnosis of ovarian cancer must consider the risk of the diagnosis being a false positive, and the possible consequences of surgery. Understanding the notion of false positives, false negatives, as well as true positives and true negatives. These 4 possible outcomes are probabilistic and not certainties in most situations (conceptual understanding).
Statistical Inference	Inference is the act of choosing an interpretation from observations. Statistical inference is based on observations sampled from a population, which are the result of a random phenomenon calculated through various statistical tests mediated by a model of the population. These tests are chosen according to the type of observations, the study design and the research question (conceptual understanding).

Notes. The procedural skills are in bold.

3.4 Lack of Consensus around Correct Responses

As a complementary analysis, the participants' scores were examined. The participants were informed that the focus was on their assessments regarding the concepts, but many nonetheless answered the items.

Average scores are relatively low, ranging from 66% in the SRA to 83% in the SRBCI. Corresponding tables (Tables S9–S12) are available on OSF at https://osf.io/t6w7n/. These scores represent the average percentage of participants' score. Remember that the participants are all statistics instructors. It indicates that the tests are not as clear as one would assume. Indeed, there does not seem to be a consensus on what the correct answer for each item is among experts. This can be further examined in participants' text entries presented. Also, note that the best-succeeded questionnaire (SRBCI) was built for biology students, whereas the recruited participants are researchers and instructors in the psychological sciences. Hence, the low scores in the other three tests cannot be attributed to a lack of competency from the participants.

For some items, there does not seem to be an agreement between participants and the tests' developers on what the correct answer is. Some participants said that there was no single correct answer for an item. An example of such comment concerns items one to four of the SRBCI (note that all comments below come directly from the participants' responses):

> "Given some of the wording there is no set answer to some of the questions, so it depends on how a student was educated. Not having the test in front of me I can't recall exactly the issues was one where there was an increasing effect, and the question is whether that increasing effect was significant (I think). A rational look would argue that taking all of the CIs into effect and the overall increasing effect would argue that you have something there. But some people, under a stricter hypothesis testing (and multiple testing)

education that may even defy logic at times would focus on one of the increases being large enough that CIs don't overlap […]"

Other participants argue that the correct answer is none of the response choices given. One participant said that item 4 of the SRA does not have the actual correct answer in its choices:

"This could be seen as a trick question, as the use of the word 'typical' hints that the mode is being sought. […] Frankly, given that outlie[r] of 22, I'd be more inclined to choose the median as my measure of central tendency but that doesn't appear as an option."

There also seems to be confusion in the participants' responses to some items. Participants do not seem sure which response choice they would choose. Some argue that the possible responses do not match the theory behind a concept. For example, one said for item 22 of the CAOS:

"I'm guessing maybe you wanted the first answer. None of the others are as good. But even though the design doesn't permit causal conclusions one can sometimes draw logical conclusions that are reasonable. One can conclude that earning more money does cause more recycling because the opposite is ridiculous. What one cannot do is draw the strong conclusion that the cause is direct and psychological. So, the alternative presented here with a conclusion is clearly wrong because it goes after a deeper direct cause that's completely unwarranted. This requires a surface understanding that correlation does not mean causation."

This spontaneous discussion of correlation vs. causation shows that despite the validation and reliability testing done with CAOS, the wording of the items is not ideal.

4 Discussion

The goal of the present study was to understand what is evaluated by the four standardized statistics understanding tests. A subsidiary goal was to examine participants' scores on those tests. In other words, the objective was to examine, from the perspective of experts, which statistical concepts are measured and whether experts agree with the tests' developers on the correct answers. The codes—representing the concepts participants listed—of each item were extracted, and then the themes—representing the concepts measured by the tests—were synthesized into a list of basic statistical concepts given in Table 2. Those concepts may be thought to be the building blocks of statistics, at least if these four tests are *adequate* and *exhaustive* reflections of statistics *understanding*.

Three intermediate conclusions can be reached regarding these assumptions. (i) Some participants expressed concerns that the material used for assessments may not be adequate. They suggested that the items may use biased formulation or skewed options. Their low scores on the tests (in particular for the SRA with only two-thirds of correct answers) lend support for this critique. Using four distinct tests hopefully alleviated this potential problem. (ii) The SRBCI encompassed only three statistical concepts, an incredibly small number given that a total of 19 (plus one miscellaneous) themes were identified. Whether such tests lack exhaustiveness or the themes given are “high-level” concepts which hide multiple drawer concepts, so to speak, remains to be clarified. (iii) The results suggest that memorization of concepts plays a critical role in the tests. For example, many items rely on remembering what a sample is (its definition) but not necessarily the implications or consequences of *varying* the sample sizes. Likewise, many assumed that the short test's names were known and remembered (*t* test, *z* test, chi-square test) which does not measure the understanding of the inner cogs of these procedures, only the ability to recognize their labels. As such, they may measure statistical literacy more than statistical understanding, or they may measure more procedural skills than conceptual understanding.

4.1 Methodological Limitations

As pertinent as the results may be, it is necessary to consider the limitations of the present study. First, the sample size was smaller than expected. Most instructors who clicked on the link did not finish the questionnaire, and so were excluded from analysis. Small samples have some disadvantages. Specifically, the sample size may have affected the proportions and diversity of codes and themes that were identified. Indeed, with a larger sample, the list of fundamental statistical concepts may be different. The small sample size may have affected the randomization of blocks of item(s). When looking at the number of responses for each block, we see that some blocks of item(s) have more responses than others. For example, the block of items 3 to 5 of the CAOS has 14 responses, but its block of item 16 has only eight. This unequal distribution may have affected the codes' recurrence, their proportion, the themes identified, and participants' mean score. Also, the small number of responses per item (between 6 and 15) could have affected the qualitative and quantitative results. A larger sample would have minimized these issues. Relatedly, a single rater did the thematic analysis. Considering how subjective the analysis is, this may have influenced the results. For example, a theme may have included different codes, and additional codes or themes may have been found.

Another limitation concerns the target population. It is believed that the chosen sample was better suited to participate in the present study for the reasons mentioned in the Methods section. However, the less experienced participants in teaching statistics may have provided more incorrect answers to the tests' items. Also, the perspective of two other populations could have provided more balanced data. Indeed, students and administrators could have provided interesting responses, which could have influenced the results. It is possible that additional codes and themes could have been identified, or the recurrence of the themes could have been influenced, which would modify the current list of basic statistical concepts.

A last limitation concerns the tests chosen as the study material. It was acknowledged that there are other tests measuring statistics understanding, and while these did not fit with the study's objectives, there is one that could have been added. This test is the LOCUS project. However, the test was developed based on the GAISE recommendations for PreK-12 statistics education within the mathematics curriculum.[21] Considering that the objective of the present study was to investigate tests measuring statistics understanding of college or university students in psychology, the LOCUS was not used. Also, while the same statistical concepts may be taught across disciplines, it is necessary to keep in mind that these tests were developed for different populations; the SRBCI, for example, was developed for biology students.

4.2 Theoretical Limitations

Naming the statistical concepts that must be learned will allow the development of assessment tools that are valid with respect to content, which in turn will allow proper evaluation of students' understanding. The approach taken herein makes the overarching assumptions that there are distinct concepts in statistics that can be isolated. However, this assumption could be wrong. Maybe statistical concepts are an intrinsically intertwined web of interrelated concepts. Take the confidence interval for example. To understand that this concept provides a range of plausible values for a descriptive statistic with a certain risk level, one needs to understand sampling, central tendency, variability, and probability. Yet, to figure out variability, a range of values is an apt representation which is precisely the desired understanding to reach from CIs. In other words, the concepts could be like Kekulé's snake biting its own tail. As such, there may not be one ideal teaching sequence for statistics, or one ideal curriculum ordered in an ideal way. As another example, some quantities are meta-statistics: the standard error is the standard deviation of descriptive statistics, that is, a measure of spread of a second order. Thus, the concept of spread has multiple layers of existence. It may be that repetition is the key to teaching statistics. By being taught the concepts and their use multiple times, each additional iteration presenting more abstract applications, it may become possible to gain a deeper understanding of those concepts. In other words, maybe learning the same concepts multiple times is a necessity in statistics education to create a conceptual map that gets bigger, clearer, and more nuanced every time.

Another assumption adopted herein is the view that statistical concepts embody static knowledge. In this view, a concept is a fixed declaration, like, for example, “2 + 2 = 4”. Maybe this assumption is wrong as well. Statistical concepts could be operations that make elements move into new states. For example, sampling could be seen as the displacement of a few objects from a general pool to a smaller pool, called “the sample” (Stuart[53] made a similar argument for *population*). Thus, what would matter in statistics teaching is not the transmission of fixed concepts, but the transmission of transformations. Students may not be well equipped (in terms of cognitive abilities) to manipulate statistics components. Consequently, teaching how to mentally manipulate elements could facilitate the understanding of statistics.[27] To investigate this possibility, one may look at the contribution of mental manipulation abilities to statistics understanding (a study currently in preparation). In any case, a better questionnaire measuring conceptual understanding of statistics instead of memorization and procedural skill capacities is needed.

4.3 Implications

This list of basic statistical concepts (seen in Table 2) has implications for statistics education. As mentioned in the introduction, there are many issues in the education of statistics, and many studies argue that the curriculum is not well developed. Due to advancements in technology, many no longer recommend teaching (and assessing) students (on) how to compute a statistic by hand (e.g.,[12]). In other words, many no longer recommend teaching statistical procedural skills. Also, some research argues that there are missing or poorly described concepts in the curriculum, such as the concept of variability.[42] The present study is a starting point toward the improvement of the teaching of statistics through conceptual understanding. With an in-depth examination of four tests on statistics understanding, it was possible to identify specific statistical concepts that are being measured in these tests. Those concepts can be considered as building blocks of statistics and organized with prerequisites. For example, to understand the concept of variability, one needs to understand the concept of population and sample/sampling first. As such, an instructor can first teach the concept of population, then of sample/sampling, and then of variability with different methods to measure it (e.g., CIs, standard deviation, etc.).

Having the building blocks of statistics—what must be taught to students and so assessed—would solve many issues in statistics education. The list of basic concepts provides insights into what is or should be typically taught in introductory statistics courses. In essence, it provides professors and educators with a curriculum. In turn, knowing what is taught gives us insights into the learning objectives of those courses. By knowing what to teach, instructors will be better prepared and will be able to build better assessments and assignments that will support their students' learning, fostering deeper conceptual understanding. Furthermore, it may stimulate research into statistics education.

Finally, the qualitative results of the present study indicate the need for better developed instruments that measure students' conceptual understanding rather than memorization capacity and procedural skills. Indeed, some themes identified did not match the concepts the tests were designed to measure. When looking at the quantitative data, the results suggest that experts did not agree on the correct answers. Some even argued what the correct answer should be and why. This could either mean that statistics instructors in social sciences are not as good at statistics and hold some misconceptions that they transfer to students, or that the tests are not as clear as first thought. Both explanations have big implications for the education of statistics. The first explanation brings to light the difficulty and non-intuitive nature of statistics, which is not new. What is novel, however, is that these characteristics persist even after many years of instruction and experience. This may point to the need for further research on professors' statistical ability. For example, it would be interesting to do the same study with statistics professors coming from statistics departments and compare the results. Likewise, it would be interesting to investigate if there is a discrepancy between what statisticians say the learning goals should be and what is taught in classes, as a reviewer of the present text pointed out. Another possible study is to examine if it is an issue with the training of non-statistician statistics instructors, like psychology professors. The second explanation supports the conclusion drawn from the qualitative results, which points to more research on assessments in statistics.

4.4 Conclusion

Through an investigation of four standardized tests on statistics understanding, it was possible to identify basic concepts that are typically taught to university students. Statistics instructors were asked to identify the concepts measured in each item. They also had the possibility of answering the items. The results suggested that there are 19 concepts measured in those four tests. Also, the instructors that answered the items had scores of 69% for the SRA, 80% for the SCI, 75% for the CAOS, and 83% for the SRBCI, for a grand mean of 76.8%. The weighted average is 76%. This large variance across tests calls for additional research on devising statistics assessments. In the end, this research may benefit the teaching and learning of statistics to foster understanding over memorization, form more efficient citizens, and future instructors to teach future students.

ACKNOWLEDGMENTS

This research began while the first author was completing a Practicum in Research taught by Jean-François Bureau at the University of Ottawa during the 2021–2022 school year. She thanks him for the opportunity and the classmate who reviewed an earlier version anonymously.

FUNDING INFORMATION

This work was supported by the Social Sciences and Humanities Research Council under Grant number RGPIN-2024-03733.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data is available on OSF (https://osf.io/t6w7n/).

References

1. R. Ackerman, D. M. Bernstein, and R. Kumar, Metacognitive hindsight bias, Mem. Cognit. **48** (2020), no. 5, 731-744. https://doi.org/10.3758/s13421-020-01012-w.
2. K. Allen. The statistics concept inventory: Development and analysis of a cognitive assessment instrument in statistics, Doctoral dissertation, University of Oklahoma. SSRN Journal. 2006 https://doi.org/10.2139/ssrn.2130143.
3. F. Bakar and V. Kumar, The use of humour in teaching and learning in higher education classrooms: Lecturers' perspectives, J. Engl. Acad. Purp. **40** (2019), 15-25. https://doi.org/10.1016/j.jeap.2019.04.006.
4. S. Béland, D. Cousineau, and N. Loye, Utiliser le Coefficient Omega de McDonald à la Place de l'Alpha de Cronbach, McGill J. Educ. **52** (2017), no. 3, 791-804. https://doi.org/10.7202/ 1050915ar.
5. J. Benson, Structural components of statistical test anxiety in adults: An exploratory model, J. Exp. Educ. 57 (1989), no. 3, 247-261. https://doi.org/10.1080/00220973.1989.10806509.
6. M. Birenbaum and S. Eylath, Who is afraid of statistics? Correlates of statistics anxiety among students of educational sciences, Educ. Res. **36** (1994), no. 1, 93-98. https://doi.org/10.1080/0013188940360110.
7. W. J. Boone, Rasch analysis for instrument development: Why, when, and how? CBE Life Sci. Educ. **15** (2016), no. 4, 1-7. https://doi.org/10.1187/cbe.16-04-0148.
8. V. Braun and V. Clarke, "Thematic analysis," APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological, H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher (eds.), American Psychological Association, Washington, DC, 2012, pp. 57-71.

9. P. Burckhardt, R. Nugent, and C. R. Genovese, Teaching statistical concepts and modern data analysis with a computing integrated learning environment, J. Stat. Data Sci. Educ. **29** (2021), no. sup1, S61-S73. https://doi.org/10.1080/10691898.2020.1854637.

10. M. Cantinotti, D. Lalande, M.-A. Ferlatte, and D. Cousineau, Validation de la Version Francophone du Questionnaire d'Anxiété Statistique (SAS-F-24), Can. J. Behav. Sci. **49** (2017), no. 2, 133-142. https://doi.org/10.1037/cbs0000074.

11. S. K. Carpenter, Testing enhances the transfer of learning, Curr. Dir. Psychol. Sci. **21** (2012), no. 5, 279-283. https://doi.org/10.1177/0963721412452728.

12. R. Carver, M. Everson, J. Gabrosek, N. Horton, R. Lock, M. Mocko, A. Rossman, G. H. Roswell, P. Velleman, J. Witmer, and B. Wood, Guidelines for assessment and instruction in statistics education (GAISE). College Report 2016, American Statistical Association, Alexandria, VA, 2016.

13. B. L. Chance, Components of statistical thinking and implications for instruction and assessment, J. Stat. Educ. **10** (2002), no. 3, 1-14. https://doi.org/10.1080/10691898.2002.11910677.

14. K. M. T. Collins and A. J. Onwuegbuzie, I cannot read my statistics textbook: The relationship between reading ability and statistics anxiety, J. Negro Educ. **76** (2007), no. 2, 118-129. https://www.jstor.org/stable/40034551.

15. D. Cousineau and B. Harding, Pourquoi les Statistiques sont-elles Difficiles à Enseigner et à Comprendre? Quelques Réflexions, Rev. Psychoeduc. **46** (2017), no. 2, 397-419. https://doi.org/10.7202/1042257ar.

16. L. J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika **16** (1951), no. 3, 297-334. https://doi.org/10.1007/BF02310555.

17. T. Deane, K. Nomme, E. Jeffery, C. Pollock, and G. Birol, Development of the statistical reasoning in biology concept inventory (SRBCI), CBE Life Sci. Educ. **15** (2016), 1-13. https://doi.org/10.1187/cbe.15-06-0131.

18. R. DelMas, J. Garfield, A. Ooms, and B. Chance, Assessing students' conceptual understanding after a first course in statistics, Stat. Educ. Res. J. **6** (2007), no. 2, 28-58. http://www.stat.auckland.ac.nz/serj.

19. R. DelMas, Statistical literacy, reasoning, and thinking: A commentary, J. Stat. Educ. **10** (2002), no. 2, 1-11. https://doi.org/10.1080/10691898.2002.11910674.

20. T. A. DeVaney, Anxiety and attitude of graduate students in on-campus vs. online statistics courses, J. Stat. Educ. **18** (2010), no. 1, 1-15. https://doi.org/10.1080/10691898.2010.11889472.

21. C. Franklin, G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer, Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-K-12 curriculum framework, American Statistical Association, Alexandria, VA, 2007.

22. J. Garfield and B. Chance, Assessment in statistics education: Issues and challenges, Math. Think. Learn. **2** (2000), no. 1–2, 99-125. https://doi.org/10.1207/S15327833MTL0202_5.

23. J. Garfield and I. Gal, "Teaching and assessing statistical reasoning," Developing mathematical reasoning in grades K-12, L. Stiff and F. R. Curcio (eds.), National Council of Teachers of Mathematics, Reston, VA, 1999, pp. 1-17.

24. J. Garfield, Beyond testing and grading: Using assessment to improve student learning, J. Stat. Educ. **2** (1994), no. 1, 1–10. https://doi.org/10.1080/10691898.1994.11910462.

25. J. Garfield, Assessing statistical reasoning, Stat. Educ. Res. J. **2** (2003), no. 1, 22-38. https://doi.org/10.52041/serj.v2i1.557.

26. J. Garfield and D. Ben-Zvi, Developing students' statistical reasoning: Connecting research and teaching practice, Springer, New York, 2008.

27. R.-M. Gibeau, E. A. Maloney, S. Béland, D. Lalande, M. Cantinotti, A. Williot, L. Chanquoy, J. Simon, M.-A. BoislardPépin, and D. Cousineau, The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender, J. Numer. Cogn. **9** (2023), no. 1, 16-43. https://doi.org/10.5964/jnc.8199.

28. G. Gigerenzer, How to make cognitive illusions disappear: Beyond "heuristics and biases", Eur. Rev. Soc. Psychol. **2** (1991), no. 1, 83-115. https://doi.org/10.1080/14792779143000033.

29. H. Harth. An analysis of the teaching of introductory statistics at university in "context," Doctoral dissertation, Loughborough University. 2018.

30. P. J. Holmes. *The effect of motivation and attitude towards statistics on conceptual understanding of statistics*, Doctoral dissertation, University of Georgia. 2014.

31. R. Hubbard, Assessment and the process of learning statistics, *J. Stat. Educ.* **5** (1997), no. 1, 1–9. https://doi.org/10.1080/10691898.1997.11910522.

32. T. Husereau, D. Cousineau, S. Zang, and R.-M. Gibeau, A compendium of common heuristics, misconceptions, and biased reasoning used in statistical thinking, Quant. Methods Psychol. **20** (2024), no. 1, 57-75. https://doi.org/10.20982/tqmp.20.1.p057.

33. J. L. Jensen, M. A. McDaniel, S. M. Woodard, and T. A. Kummer, Teaching to the test ... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding, Educ. Psychol. Rev. **26** (2014), no. 2, 307-329. https://doi.org/10.1007/s10648-013-9248-9.

34. X. Jia, W. Hu, F. Cai, H. Wang, J. Li, M. A. Runco, and Y. Chen, The influence of teaching

methods on creative problem finding, Think. Skills Creat. **24** (2017), 86-94. https://doi.org/10.1016/j.tsc.2017.02.006.

35. D. Kahneman and A. Tversky, Subjective probability: A judgment of representativeness, Cogn. Psychol. **3** (1972), no. 3, 430-454. https://doi.org/10.1016/0010-0285(72)90016-3.
36. R. Konicek-Moran and P. Keeley, Teaching for conceptual understanding in science, National Science Teachers Association, Richmond, VA, 2015.
37. C. Konold, Informal conceptions of probability, Cogn. Instr. **6** (1989), no. 1, 59-98. https://doi.org/10.1207/s1532690xci0601_3.
38. C. Konold, Issues in assessing conceptual understanding in probability and statistics, J. Stat. Educ. 3 (1995), no. 1, 1-9. https://doi.org/10.1080/10691898.1995.11910479.
39. S. M. Marson, Three empirical strategies for teaching statistics, J. Teach. Soc. Work. **27** (2007), no. 3-4, 199-213. https://doi.org/10.1300/J067v27n03_13.
40. M. M. McIntyre, Statistics is puzzling: Testing a novel approach to statistics learning, Scholarsh. Teach. Learn. Psychol. **9** (2020), no. 2, 1-9. https://doi.org/10.1037/stl0000204.
41. P. Mcmahon, T. Zhang, and R. Dwight, Requirements for big data adoption for railway asset management, IEEE Access **8** (2020), 15543-15564. https://doi.org/10.1109/ACCESS.2020.2967436.
42. M. M. Meletiou. Developing students' conceptions of variation: An untapped well in statistical reasoning, Doctoral dissertation, University of Texas. 2000.
43. D. Nolan and D. Temple Lang, Computing in the statistics curricula, Am. Stat. **64** (2010), no. 2, 97-107. https://doi.org/10.1198/tast.2010.09132.
44. A. J. Onwuegbuzie and M. A. Seaman, The effect of time constraints and statistics test anxiety on test performance in a statistics course, J. Exp. Educ. **63** (1995), no. 2, 115-124.

https://doi.org/10.1080/00220973.1995.9943816.

45. R. Park, Practical teaching strategies for hypothesis testing, Am. Stat. **73** (2019), no. 3, 282-287. https://doi.org/10.1080/00031305.2018.1424034.
46. G. Rasch, Probabilistic models for some intelligence and attainment tests, MESA Press, San Diego, CA, 1993.
47. B. Rittle-Johnson, R. S. Siegler, and M. W. Alibali, Developing conceptual understanding and procedural skill in mathematics: An iterative process, J. Educ. Psychol. **93** (2001), no. 2, 346-362. https://doi.org/10.1037/0022-0663.93.2.346.
48. A. Robitzsch, T. Kiefer, and M. Wu. TAM: Test analysis modules [R package version 4.2-21]. 2024 https://CRAN.R-project.org/package=TAM.
49. A. Sabbag, J. Garfield, and A. Zieffler, Assessing statistical literacy and statistical reasoning: The REALI instrument, Stat. Educ. Res. J. **17** (2018), no. 2, 141-160. https://doi.org/10.52041/serj.v17i2.163.
50. S. Sharma, Definitions and models of statistical literacy: A literature review, Open Rev. Educ. Res. **4** (2017), no. 1, 118-133. https://doi.org/10.1080/23265507.2017.1354313.
51. M. J. Shaughnessy, Statistics for all—The flip side of quantitative reasoning, National Council of Teachers of Mathematics, Reston, VA, 2010. https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/J_-Michael-Shaughnessy/Statistics-for-All—the-Flip-Side-of-Quantitative-Reasoning/.
52. K. Slootmaeckers, B. Kerremans, and J. Adriaensen, Too afraid to learn: Attitudes towards statistics as a barrier to learning statistics and to acquiring quantitative skills, Politics **34** (2014), no. 2, 191-200. https://doi.org/10.1111/1467-9256.12042.
53. M. Stuart, Changing the teaching of statistics, J. R. Stat. Soc. Ser. D Stat. **44** (1995), no. 1,

45-54. https://doi.org/10.2307/2348615.

54. L. Theis and A. Savard, Linking probability to real-world situations: How do teachers make use of the mathematical potential of simulation programs? [Contributed paper refereed], Eighth International Conference on Teaching Statistics (ICOTS8), Ljubljana, Slovenia, 2010.
55. A. Tversky and D. Kahneman, Judgment under uncertainty: Heuristics and biases, Science **185** (1974), no. 4157, 1124-1131. https://www.jstor.org/stable/1738360.
56. A. Vigil-Colet, U. Lorenzo-Seva, and L. Condon, Development and validation of the statistical anxiety scale, Psicothema **20** (2008), no. 1, 174-180. https://reunido.uniovi.es/index.php/PST/article/view/8638.
57. A. F. Villarejo-Ramos, J.-P. Cabrera-Sanchez, J. Lara-Rubio, and F. Liébana-Cabanillas, Predicting big data adoption in companies with an explanatory and predictive model, Front. Psychol. **12** (2021), 1-12. https://doi.org/10.3389/fpsyg.2021.651398.
58. J. Watson and R. Callingham, Statistical literacy: A complex hierarchical construct, Stat. Educ. Res. J. **2** (2003), no. 2, 3-46. https://doi.org/10.52041/serj.v2i2.553.
59. M. C. Wittmann, J. T. Morgan, and R. E. Feeley, Laboratory tutorial activities for teaching probability, Phys. Rev. Spec. Top. Phys. Educ. Res. **2** (2006), 1-8. https://doi.org/10.1103/PhysRevSTPER.2.020104.

Chapter 5: A Fourth Tentative Study

Study D

Development of the Yet Another Conceptual Understanding of Statistics Assessment (YACUSA)

Disclaimer

This was expected to be the fourth and last study of the present thesis. Originally, Study A and B were prerequisite studies, which were, along with Study C of the present thesis, leading up to this last study. However, despite our best efforts and our many attempts, the development of the YACUSA was not entirely successful. This fourth study is still presented for transparency. Further, we believe that the limited success of this project might help understand the faulty assumptions that were part of its initial conception.

Abstract

One big issue in statistics education concerns the evaluation of student learning. There have been some recommendations over the years, but none that concretely help instructors to assess their students' *conceptual* understanding. The present study investigates the reliability and validity of the *Yet Another Conceptual Understanding of Statistics Assessment* (YACUSA). We developed this questionnaire to measure understanding of statistics in university students without relying on memorization and step-by-step operations. A total of 368 participants were recruited. The results indicate that the YACUSA has low reliability (measured with intra-class correlation) and may be best structured as a three-factor model, which are defined as three different skills developed in statistics courses. Other results show that the items have good discriminability between ability levels and that they are mostly easy. Also, the total score on YACUSA correlates well with other variables (self-reported statistics grade, statistics anxiety and spatial anxiety). It is possible that statistics may be best understood as a series of skills instead of a series of concepts, which may explain the unexpected results. Nevertheless, this assessment tool could be a great addition to statistics education as it opens doors to new research.

Keywords: Statistics education, assessment, conceptual understanding, standardized statistics test

Development of the *Yet Another Conceptual Understanding of Statistics Assessment* (YACUSA)

The teaching of statistics is challenging at any level of education. The challenge is present right from a class inception when educators are developing the curriculum. Indeed, it is not clear what instructors should teach and in what order (Carver et al., 2016; Garfield & Ben-Zvi, 2007; Shaughnessy, 2019). Then, the challenge is encountered again when instructors teach, as teaching methods influence students' learning and performance (e.g., DeVaney, 2010; Harth, 2018). Finally, the challenge is encountered when instructors assess student learning (e.g., Chance, 2002). These challenges have been studied and discussed since the beginning of statistics education, dating back to the late 1800s (e.g., Chaddock, 1926; Cobb, 1992; Franklin et al., 2007; Snedecor, 1948; Tucker et al., 1959; Zavez & Harel, 2025; Zieffler et al., 2018). Yet they are still unresolved.

Research on Statistics Education

Debates on statistics education are not recent. In the late 1800s and early 1900s, the discussion was centered around the implementation of statistics courses in universities and colleges (Zieffler et al., 2018). Statistics was taught as a tool for research. However, some researchers argued that this instruction is not sufficient (Chaddock, 1926). According to them, instructors should instead strive to develop statistical reasoning in students. In the 1930s, researchers saw the rise of *Mathematical statistics* as a subject (David, 1998). Around the same time, courses in statistics started to include more advanced methods, like correlation and analysis of variance (Zieffler et al., 2018). After World War II, the necessity of trained statisticians and lack thereof were particularly present. During this time, research on statistics education started, with researchers investigating the concepts to teach and the optimal curriculum sequence.

University programs in statistics then started to be implemented. In the 1950-60s, recommendations about teaching statistics and probability in high school were made (Tucker et al., 1959).

These debates brought other questions, like how to teach, how students learn statistics and what impedes their understanding (e.g., Kahneman & Tversky, 1972). Some researchers observed that students have many misconceptions, which impact their ability to learn and understand statistics and probability (Tversky & Kahneman, 1974). In other words, students are not good when facing uncertainty (some would even argue that humankind as a whole is not good with uncertainty; Béland et al., 2016; Gibeau et al., 2023; Kahneman, 2011). Indeed, Kahneman, Tversky and their colleagues found that students hold many mistaken reasonings: for two examples, the *Representativeness Heuristic* and the *Availability Heuristic* (Konold, 1995; Tversky & Kahneman, 1974; see Husereau et al., 2024, for a review of misconceptions).

Another factor impeding understanding of the topic that started to emerge in the 1980s and 1990s is *statistics anxiety*. Many researchers observed that most students feel anxious toward statistics and that this anxiety influences their learning and performance (e.g., Benson, 1989; Birenbaum & Eylath, 1994; Cruise et al., 1985). Furthermore, it was observed that statistics anxiety is the best predictor of the students' grade (Cantinotti et al., 2017; Chiesi & Primi, 2010; Fullerton & Umphrey, 2016; see Mackinnon et al., 2025, for a counter-evidence). Note however that Studies A and B of Chapters 2 and 3 of the present thesis suggest that this anxiety toward statistics is predicted by spatial anxiety. Those issues of uncertain curriculum, inefficient teaching methods, students' incorrect reasonings and anxiety raised another question: what to evaluate and how.

The Challenge of Interest

Since the early 2000's, researchers have investigated how statistics instructors can assess learning. This issue resulted in a few tests, some of which claiming to measure conceptual understanding (Allen, 2006; Deane et al., 2016; DelMas et al., 2007; Garfield, 2003). An example item taken from Allen's (2006) dissertation is shown in Figure 1. This item seems to assess students' capacity to memorize information, which is very different from students' capacity to conceptually understand what they learned. Further, many items of these tests require procedural skills (i.e., applying formulas or statistical procedures; the definition is taken from Rittle-Johnson et al., 2001) and memorisation (Gibeau & Cousineau, 2025, Study C of Chapter 4 of the present thesis).

The distinction between conceptual understanding and procedural skills is essential in a field like statistics, where concepts are abstract and intertwined (Gibeau & Cousineau, 2025, Study C of Chapter 4 of the present thesis; Bargagliotti, et al., 2020). Considering that the goal of statistics courses is to train statistically literate citizens who understand societal debates and who help in maintaining and advancing our society, teaching and evaluating procedural skills do not help students achieve this goal. Indeed, to do so, they need to understand the nuances of statistics; the signification of statistical concepts, and how they influence each other. This *conceptual understanding* was defined as the "understanding of the representation of similarities and differences (or characteristics) of elements and the association between them or how they influence each other [43]" (Gibeau & Cousineau, 2025, p.4).

Even when a test is developed to measure conceptual understanding, there are still many technical terms, procedures and memorization involved. If such a test still depends on procedural knowledge despite researchers' effort, how can one measure true conceptual understanding?

A Potential Solution

One possibility is to develop a test without technical terms, without formulas and as little numbers as possible. Doing this will allow instructors to evaluate students' understanding of concepts' signification and implications.

To develop such a test, one first needs to know what to measure. Ideally, this test would measure all basic statistical concepts (i.e., concepts taught in introductory courses). However, those concepts vary depending on the institution and on the instructor. For example, Nolan and Temple Lang (2010) suggest three main sections: "*Statistical computing*" (p. 98), "*computational reasoning and literacy*" and then applying statistical computations to real-world problems. Other researchers propose to organize the curriculum by variable types or to teach by cases (Malone et al., 2010). Specifically, instructors start by teaching basic definitions, sampling methods and study designs. Then, they teach analyses with (one and multiple) categorical variables, and move on to analyses with (one, one in different groups and multiple) numerical variables. When teaching by cases instructors repeatedly show the statistical analysis process in different contexts. For example, when teaching one-sample quantitative data, the instructor first show graphical representation, teach how to run descriptive statistics, and how to analyse these types of data.

As can be seen, the curricula in the above examples are very different. As such, one cannot rely on them to know what should be included in a test measuring conceptual understanding. In a recent qualitative study, the authors relied on existing tests (Gibeau & Cousineau, 2025, Study C of Chapter 4 of the present thesis). They gathered four instruments claiming to measure conceptual understanding of statistics in college or university students. These tests are the following: the *Statistical Reasoning Assessment* (SRA; Garfield, 2003), the

Statistics Concept Inventory (SCI; Allen, 2006), the *Comprehensive Assessment of Outcomes for a First Course in Statistics* (CAOS; DelMas et al., 2007), and the *Statistical Reasoning in Biology Concept Inventory* (SRBCI; Deane et al., 2016). Then, they asked statistics instructors which concept(s) each item measure. The study resulted in a list of 18 statistical concepts. This list is presented in Study C of the present thesis.

The aim of the present study is to develop a novel test measuring students' understanding of statistical concepts found in Gibeau and Cousineau (2025). This test will try to capture students' conceptual understanding by removing all formula, technical terms and by having as little numbers as possible. In addition, because performance in statistics is predicted by statistics anxiety and that this type of anxiety is predicted by spatial anxiety, these last two variables will be measured and used as controls.

Method

Prior to data collection, this study was reviewed and approved by the *Ethics Research Board* of the authors' affiliation. We decided to only present iteration six in the present text because this version is better than the others. For transparency, the first five iterations are presented in Appendix B of the present thesis.

In iteration one, 53 items divided into 16 concepts (described in Study C of the present thesis) were tested. This version's internal consistency was good. The factor structure was weak as many items loaded on unexpected factor or did not load on any factor. The Item-Response Theory (IRT) models revealed that a little more than half of the items differentiated well between all ability levels and that a little less than half were too easy. Considering these results, we decided to reduce the number of concepts (subscales) measured in iteration two. We kept only the concepts that were almost certainly taught in all introductory statistics courses (i.e., population, randomness & probability, sample, variability & dispersion, and central tendencies). This version had 17 items divided in five subscales. The internal consistency and factor structure were weak, and most items were too easy. However, about half of the items differentiated well between all ability levels. We removed problematic items, created some items and improved the ones that were kept for iteration three. Also, the "population" and "sample" concepts were assembled as it is believed that one cannot go without the other. This version had 34 items divided in four subscales. The internal consistency was good, and the factor structure was acceptable at best. The IRT models were similar to iteration two. We removed problematic items and created or improved the items for iteration four. A similar pattern of results was found. For iteration five, some items were removed, some were created and others were improved. A fifth concept was added to the test (*Shared Variance & Correlation*) as it is believed that understanding this concept is necessary to understand more advanced ones. This version had 26 items divided into five concepts. The internal consistency was acceptable-to-good, and the factor structure was weak-to-acceptable. The IRT models revealed that most items differentiated well between all ability levels and that they were too easy. For the good of the thesis, it was decided to not investigate this matter further. Instead, a replication study of iteration five is presented herein.

Participants

Participants were recruited through the *Integrated System of Participation to Research* (ISPR) of the authors' affiliation during the Fall semester of 2025. This system allows professors and graduate students to make their study available to undergraduate students. These undergraduate students participate in studies in exchange for course credits in their introduction to psychology courses. Students were told that this study is about a statistics understanding test and that their responses would be used to improve it. They were told that there was no formula and that they only had to answer to the best for their ability. The target sample size was 500 based on previous studies (e.g., Lyons et al., 2018). However, only 453 students clicked on the link to the questionnaire.

Most participants identified as women (75% of the 368 participants retained for analysis), many identified as man (23%), and a few identified as non-binary (0.01%). Other participants chose "Prefer not to say" or "Other" (0.01%). About half of the participants identified as Caucasian or White (49%) and Asian (and Middle East) or Pacific Islander (29%). Some participants identified as African American or Black (6%) and North African or Arab (5%). A few participants identified as Mixed heritage (4%), Hispanic or Latinx (2%) and Indigenous, First Nations or Metis (1%). Almost all participants were between 18-24 years old (approximately 98%), some were between 25-34 years old (2%) and between 35-44 years old (1%). Most participants were studying psychology (21%), health sciences (17%), and biomedical science (12%). Some were studying nursing (8%), business (7%), accounting (5%), criminology (5%), and kinetics (3%). A few participants were studying political science (2%), marketing (2%), management (1%), linguistics (1%), communication (1%), sociology (1%), engineering (1%), computer science (1%), biology (1%), biochemistry (1%), social work (0.08%), anthropology (0.08%), arts (0.08%), philosophy (0.05%), law (0.05%), mathematics (0.03%), journalism (0.03%) and economics (0.03%). Other participants selected "Other" (5%) and "Please, select a choice" (0.03%).

Material

All items were inspired by previous work (e.g., Konold, 1995; Tversky & Kahneman, 1974), teaching experiences of the thesis supervisor or created *ex nihilo*. The concepts identified in Gibeau and Cousineau (2025) were used herein as YACUSA's subscales. Two additional constructs measured in the present study are statistics anxiety and spatial anxiety (described next).

YACUSA Items

A total of 13 items divided in three subscales (or concepts) were created. Table 1 presents the items. The subscales are *i*) Visual Representations (VR; items 1-4), *ii*) Randomness & Probability (RP; items 5-7), *iii*) Commitment to Interpretations (Co; items 8-13). These subscales were identified based on previous iterations. It is believed that these are basic statistical skills developed in most introductory statistics courses. In the discussion, we will discuss what our initial views on the subscales were.

Statistics Anxiety

This construct was measured by the *Statistics Anxiety Scale* (SAS; Vigil-Colet et al., 2008). This scale includes three subscales of eight items each for a total of 24 items. The subscales are Evaluation Anxiety (E; e.g., "Studying for an examination in a statistics course"), Interpretation Anxiety (I; e.g., "Interpreting the meaning of a table in a journal article"), and Anxiety of Asking for Help (A; e.g., "Going to ask my statistics teacher for individual help with material I am having difficulty understanding"). For the present study, only the Interpretation Anxiety is administered. This is because only this subscale should strongly relates to items measuring conceptual understanding of statistics (i.e., the YACUSA items). To answer the items, participants choose a response between five levels of anxiety that best applies to them. These levels range from "no anxiety" (1) to "a lot of anxiety" (5). The developers tested the scale's reliability only by looking at the Cronbach's alpha. They reported excellent internal consistency for the total scale (α = .91) and for the subscale used herein (α = .82).

Spatial Anxiety

This second construct was measure by the *Spatial Anxiety Questionnaire* (SAQ; Lyons et al., 2018). This scale has three subscales of eight items each for a total of 24 items. The subscales are Mental Manipulation (M; e.g., “Using a 3-dimensional model of an airport to complete a homework assignment”), Navigation (N; e.g., “Memorizing routes and landmarks on a map for an upcoming exam”) and Imagery (I; e.g., “Asked to recreate your favorite artist's signature from memory”). For the present study, only the Mental Manipulation and the Imagery subscales are administered. This is because a prior study indicates that only these two strongly relate to statistics interpretation anxiety (Gibeau et al., 2023). To answer the items, participants choose how much anxiety (between five levels) they would feel. The response choices range from “not at all” (1) to “a lot” (5). The developers reported that internal consistency is good to excellent (M: α = .92, and I: α = .86). To further test the scale’s reliability, the authors looked at inter-scale and inter-item correlations. All scales are moderately correlated, indicating low collinearity, and that within-scale items are more correlated than items in different scales.

Procedure and Analyses

Participants were asked to complete an online questionnaire measuring their understanding of statistical concepts (i.e., items of YACUSA), the grade they received in their last statistics course, their anxiety level when facing statistics and their anxiety level when facing activities that require spatial skills (e.g., mental rotation). Their responses and scores are used to test YACUSA’s reliability and validity. Data of iteration 6 was collected during the Fall semester of 2025.

Participants first gave their informed consent by reading the consent form and answering to the following question: “Do you wish to participate to this study?” Then, they answered

YACUSA’s items. As mentioned above, YACUSA has three subscales. These subscales were presented randomly. Next, they answered a question about their last statistics grade. After, the interpretation subscale of the SAS was presented followed by the mental manipulation and imagery subscales of the SAQ. Finally, participants answered a few sociodemographic questions (i.e., age, gender, ethnicity and program of study). This order was chosen as per Taherdoost's (2022) recommendations in question organization, where questions most related to the research question and hypotheses should come first and demographic questions should come last.

Concerning the analyses, first, basic frequencies are examined. Then, the test’s reliability is examined with intraclass correlations, Cronbach’s alpha, McDonald’s omega and factor analyses. YACUSA’s convergent validity is investigated using correlations with other measures. To further investigate the items’ reliability and validity, Item Response Theory (IRT) models were computed.

This last analysis estimates two parameters (Bean & Bowen, 2021). The alpha parameter represents the slope – whether it is abrupt or not – which concerns the differentiation capacity of the item. The more the curve is abrupt (high alpha parameter), the less the item differentiates between all ability levels. A negative alpha signifies that high-ability people have a lesser probability than others to respond correctly to the item. The beta parameter represents the threshold ability level at which one needs to be to have a 0.5 probability of responding correctly. The higher the threshold is (high beta parameter), the further away the average-ability person is from a 0.5 probability of success. The difficulty index of the item is whether the beta is negative or positive. Negative beta parameters represent easy items as people need ability less than 0 to have a 0.5 probability of success. This analysis also provides an Item Information and Characteristics Curves graphs (Bean & Bowen, 2021). This first curve shows the amount of information each item provides based on its slope. The higher the peak of the curve is, the more localized the information of the item is. The second curve shows both model parameters (the slope and the threshold), where people's probability of choosing the correct answer (Y axis) depending on their ability level (X axis). A good item should allow an average-ability person to have a 0.5 probability of choosing the correct answer (DeMars, 2010).

Results

Data Screening

A total of 453 participants clicked on the link, 413 completed the questionnaire, and 368 were kept for analysis. We removed 45 participants as they completed the questionnaire in less than 8.5 minutes, which is well below the estimated time of 17 minutes. This estimated time is suggested by Qualtrics. The system estimates the average time while considering the longest and shortest estimated time based on the number of questions, the number of words per question, and the number of response choices per question.

These 368 participants generated 13 248 responses and 42 missing responses. The Little's test was run to examine if missing values are missing at random. The test was not significant (p = .092), suggesting that the values are missing randomly. These missing responses were considered as incorrect. Among the 4 784 responses on the YACUSA items, 2 094 were correct answers and 2 690 were incorrect.

Internal Consistency

To examine YACUSA's reliability, intraclass correlation (ICC), Cronbach's alpha and McDonald's omega were used. The ICC (value of .01 with its 95% confidence interval of [−.001, .02]) is not significantly different from 0. This indicates that YACUSA does not have a good reliability. The Cronbach's alpha of the total scale (0.12) indicates that items have low internal consistency. The alpha for each of the three subscales indicates that the internal consistency is acceptable-to-good (VR: α = .39; RP: α = .63; Co: α = .43). The McDonald' omega could not be computed for the total scale nor for Co due to a lack of item covariance. However, the omega indicates acceptable-to-good reliability for the first two subscales (VR: ω = .43; RP: ω = .71).

Factor Structure

To further investigate the test's reliability, factor analyses were run. Results of the Exploratory Factor Analysis (EFA) are presented in Table 2, and the results of the Confirmatory Factor Analysis (CFA) are presented in Table 3. The 3-factor EFA model fits well with the data (RMSEA = .023, CFI = .948, TLI = .903, SRMR = .032). As can be seen in Table 2, many items do not load significantly on any factor as evidenced by the inclusion of 0 in their confidence intervals. This indicates that the factorial structure of YACUSA is weak.

A three-factor CFA analysis was run. In this analysis, the 13 items were divided according to the three factors that were defined a priori instead of being based on the EFA. The CFA has an acceptable fit with the data (RMSEA = .028, CFI = .881, TLI = .850, SRMR = .043). As seen in Table 3, only one item (i.e., item 7 onto factor 2) has a good loading estimate on its factor. The other items have an acceptable loading estimate at best (e.g., item 6), and a negative or not-significant loading at worst (e.g., item 4 and item 13, respectively). This further indicates that the a priori defined structure is weak.

Note that the EFA generates very different results compared to the CFA, which is based on the framework. Indeed, when comparing Tables 2 and 3, the same item does not load on the same factor despite that both analyses have three factors. For example, item 12 loads significantly and positively on factor 1 of the EFA, but load on factor 3 of the CFA. However, some items that loads significantly on a factor of the EFA also loads similarly on the same factor of the CFA. For example, items 1 and 2 in both analyses load similarly (in the 0.30s) on factor 1. Another example is items 6 and 7, which load similarly on factor 2 of both analyses.

Convergent Validity

To examine the YACUSA's validity, correlations were run. This is a test measuring conceptual understanding of statistics. As such, it is expected to be related to grades in past statistics courses, as well as statistics interpretation anxiety, spatial imagery anxiety, and spatial mental manipulation anxiety. Also, the Visual Representation subscale of YACUSA should better correlate with spatial mental manipulation anxiety and statistics interpretation anxiety and the Commitment to Interpretation subscale should better correlate with statistics interpretation anxiety.

Table 4 shows the correlation matrix. As can be seen, YACUSA correlated moderately with grade in statistics. A stronger correlation was expected; however, this is the self-reported grade, which is bond to some biases and errors. As expected, YACUSA has a negative moderate correlation with SAS-I, SAQ-M and SAQ-tot, as well as a non-significant correlation with SAQ-I. This last result supports previous studies which indicate that spatial mental manipulation anxiety has a stronger relation with statistics interpretation anxiety compared to spatial imagery anxiety (Gibeau et al., 2023; Gibeau et al., under review; Studies A and B of Chapter 2 and 3 of the present thesis, respectively). Also, YAC-VR is moderately negatively related to both SAQ-M and SAS-I, as expected. YAC-Co is not significantly related to SAS-I, which may indicate a validity problem, but it is related to SAQ-M. YAC-PR is not significantly related to anxiety types. Note that YAC-PR does not correlate significantly with the two other YACUSA subscales.

Item Response Theory (IRT)

To further investigate items' reliability and validity, IRT models were run with two parameters, alpha and beta. Table 5 shows the alpha (slope) and the beta (threshold) parameters. Figures 2 to 7 shows the Item Information and Characteristic Curves of each subscale. As can be seen in the Table, items 2, 3, 5 and 8 to 13 discriminate well between all ability levels. However, there seem to be a ceiling effect in most items, except for items 3, 4 and 11 to 13. The Figures seem to support these results. For example, item 4 of Figure 3 seems to discriminate between the ability levels -3 to 4 and seems a little difficult as average-ability people (0 on X axis) have an approximately 40% chance of answering correctly. Another example is item 2 of Figure 5, where the item only discriminates between the ability levels -1 to 0, and it seems too easy as average-ability people have an approximately 97% chance of answering correctly.

Discussion

In the present study, we investigated how to evaluate students understanding of the signification and implications of statistical concepts, or conceptual understanding. Initially, we used the 18 concepts identified in Study C of Chapter 4 of the present thesis (now published). Then, we focused on the five basic concepts that are typically taught in introductory statistics courses: Population & Sample, Randomness & Probability, Central tendencies, Variability & Dispersion, and Shared Variance & Correlation. Through 6 iterations of item testing, the results indicate that assessing students' conceptual understanding may not be as simple as asking them questions regarding learned course material. Indeed, the best version of the *Yet Another Conceptual Understanding of Statistics Assessment* has only three subscales, two of which are not present in Gibeau and Cousineau (2025). One includes items with graphs, which we define as students' ability to understand visual representations. A second subscale includes items on basic probability, which we defined as students' ability to understand probabilities, chance and randomness. The last subscale includes items on interpretation or committing to a certain interpretation, which we defined as students' ability to commit to an interpretation. As such, these results might lead to interesting ideas and research on statistics education, by distinguishing the influence of concepts knowledge and skills knowledge.

As it stands, the YACUSA is appropriate to use as an indicator of students' progress only. The YACUSA's reliability is acceptable at best. However, the Visual Representation and, to a lesser extent, the Commitment to Interpretation subscales do seem reliable and valid, as shown by IRT models and correlations with anxiety types and grades in previous statistics courses. Note, however, that the Commitment to Interpretation subscale did not relate significantly to statistics interpretation anxiety. This may indicate that this subscale of YACUSA does not have convergent validity. Considering the number of analyses done, this may also be a Type 1 error. The Probability & Randomness subscale does not seem valid nor reliable.

Limitations

There are some caveats that need to be discussed and can explain the lack of reliability. First, all participants were recruited from the University of Ottawa, and the majority were in social and health sciences. This means that YACUSA's items are based on students enrolled in a single university. The results (e.g., item difficulty) may only be the sample's particularities. For example, item 6 may have a ceiling effect simply because students' professors may have focused on the concept being measured. Using a sample from another university with a different concentration of programs, that item could be easier or more difficult. A sample including students in STEM programs, for example, may generate different results. A recent study indicates that social sciences students score lower on a statistics literacy measure compared to medical and economics students (Berndt et al., 2021). Relatedly, most participants were women. This might have influenced the results as many previous studies indicate a gender difference in math achievement (e.g., Leder, 1985). However, a literature review in 2020 indicate that this difference is declining over the years, and gender seems to better predict spatial skills than mathematics achievement (Wang, 2020). Note that studies investigating gender differences in statistics achievement or performance are rare. Also, asking participants to complete the study online may have influenced the results. This data collection method might have activated their spatial skills and anxiety, inflating their anxiety levels and strengthening the relation between spatial skills and anxiety as a result.

The skills measured by the items were identified based on previous iterations and on a previous study, where participants were asked to indicate what are each presented item measuring (Gibeau & Cousineau, 2025; Study C of Chapter 4 of the present thesis). The items presented to participants in that previous study were taken from four existing tests on statistics understanding. However, there are more than four that are available. If the authors investigated a different combination of existing tests, the results may have been different. Some concepts may not have come up in the data while others may have been mentioned. This means that the concepts measured in YACUSA are not exhaustive to all nor exclusive to introductory statistics courses. Indeed, some instructors may teach asymmetry or may focus on NHST, which are concepts not measured in the YACUSA. Note, however, that this instrument was specifically designed not to measure procedural skills, like NHST. Also, not all 18 concepts from Gibeau and Cousineau's (2025) study were included in YACUSA. Although items for 16 concepts were created in iteration one of the novel test, this last iteration focused on the most basic concepts, which are almost certainly taught in all introductory statistics courses. Furthermore, it is possible that using this past study based on existing tests (Gibeau & Cousineau, 2025; Study C of Chapter 4 of the present thesis) was not the correct method to develop a novel test measuring conceptual understanding of statistics. Instead, investigations on essential concepts among university administrators, educators and faculty members may be best. A third important limitation is that to develop the YACUSA, we assumed that statistics concepts can be separated from each other. However, this may not be the case and may be one of the reasons why the development of such a test did not work as expected. Indeed, it is possible that all concepts overlap significantly, so much so that they cannot be taught nor understood alone. For example, central tendencies may not be understood without considering variability and dispersion, yet these last two concepts cannot be understood without considering how they are measured (central tendencies, among others). Instructors might need to teach multiple concepts simultaneously and assess the students on concepts separately, which prevent students from seeing the complex interrelated statistical map. This blindness may make it difficult for students to truly understand statistics in a conceptual way.

Implications

Considering that the development of YACUSA did not work as expected, it is difficult to think about implications. However, negative results have signification and do advance research. Indeed, the fact that a test measuring conceptual understanding of statistics could not be developed by concept is very informative for statistics education. It is possible that such a subject cannot be separated by concept. A possible explanation for this is that those concepts are too intertwined, as said above. Another explanation is that statistics is better understood and taught as a series of skills. For example, concepts like central tendencies, variability, and distribution may all be part of students' ability to understand visual representations. Similarly, concepts like sample/sampling, population, and correlation may be part of students' ability to interpret analyses and to commit to an interpretation. This would also explain why existing tests often measure procedural skills.

This reasoning allows instructors and educators to think about a new way of teaching statistics and assessing students. Instead of assessing students' memorization capacity and procedural skills, instructors can build their course as a set of statistical skills to develop. Other than visual representations and commitment to interpretations, there could be a skill on errors and their margins where central tendencies, outliers, variability, and CIs are included. Note how central tendencies and variability are included in two different skills. This further indicates how interrelated statistical concepts are and how repetition is important.

With this way of teaching, students are forced to reason conceptually and develop their statistical skills instead of going back to textbooks for step-by-step procedures. Instead of only computing statistics, they get to build their inferential skills and see the statistics beyond their numbers. This may help in training statistically competent professionals and researchers that will push the boundaries further and contribute to societal issues and debates.

Conclusion

The *Yet Another Conceptual Understanding of Statistics Assessment* was developed and validated herein. The YACUSA was first designed to measure conceptual understanding of statistics in an unusual and innovative way to emphasize the signification and implications of statistical concepts. Instead, an unexpected structure was identified. The novel test developed, while inconclusive, opens a door for statistics education. As it is only theoretical, research needs to be done to explore the possibility that statistics are skill-based instead of concept-based.

References

Allen, K. (2006). *The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics* [Doctoral dissertation, University of Oklahoma]. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2130143

Bean, G. J., & Bowen, N. K. (2021). Item Response Theory and Confirmatory Factor Analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work*, *18*(6), 597-618. https://doi.org/10.1080/26408066.2021.1906813

Béland, S., Cousineau, D., & Loye, N. (2016). Les dix commandements du nouvel homo statisticus. *McGill Journal of Education*, *51*(2), 947-960. https://doi.org/10.7202/1038612ar

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *The Journal of Experimental Education, 57*(3), 247-261. https://doi.org/10.1080/00220973.1989.10806509

Berndt, M., Schmidt, F. M., Sailer, M., Fischer, F., Fischer, M. R., & Zottmann, J. M. (2021). Investigating statistical literacy and scientific reasoning & argumentation in medical-, social sciences-, and economics students. *Learning and Individual Differences, 86*, Article 101963. https://doi.org/10.1016/j.lindif.2020.101963

Birenbaum, M., & Eylath, S. (1994). Who is afraid of statistics? Correlates of statistics anxiety among students of educational sciences. *Educational Research, 36*(1), 93-98. https://doi.org/10.1080/0013188940360110

Cantinotti, M., Lalande, D., Ferlatte, M.-A., & Cousineau, D. (2017). Validation de la version francophone du Questionnaire d'anxiété statistique (SAS-F-24) [Validation of the French version of the Statistics Anxiety Scale (SAS-F)]. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 49*(2), 133-142. https://doi.org/10.1037/cbs0000074

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Roswell, G. H., Velleman, P., Witmer, J., & Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report 2016*. American Statistical Association. https://commons.erau.edu/publication/1083

Chaddock, R. E. (1926). The function of statistics in undergraduate training. *Journal of the American Statistical Association*, *21*(153), 1-8. https://doi.org/10.1080/01621459.1926.10502152

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, *10*(3), 1-14. https://doi.org/10.1080/10691898.2002.11910677

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal*, *9*(1), Article 1. https://doi.org/10.52041/serj.v9i1.385

Cobb, G. (1992). Teaching Statistics. In L. A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (pp.3-43). The Mathematical Association of America. https://web.archive.org/web/20180413035124id_/https://www.maa.org/sites/default/files/pdf/CUPM/pdf/CUPM_Report_1992.pdf#page=14

Cruise, R., Cash, R., & Bolton, D. (1985). Development and validation of an instrument to measure statistical anxiety. In *Annual meeting of the Statistical Education Section* (pp. 92-97). American Statistical Association.

David, H. A. (1998). Statistics in U.S. Universities in 1933 and the establishment of the statistical laboratory at Iowa state. *Statistical Science*, *13*(1), 66-74. https://doi.org/10.1214/ss/1028905974

Deane, T., Nomme, K., Jeffery, E., Pollock, C., & Birol, G. (2016). Development of the

Statistical Reasoning in Biology Concept Inventory (SRBCI). *CBE—Life Sciences Education*, *15*(1), 1-13. https://doi.org/10.1187/cbe.15-06-0131

DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28-58. http://www.stat.auckland.ac.nz/serj

DeMars, C. (2010). *Item Response Theory*. Oxford University Press.

DeVaney, T. A. (2010). Anxiety and attitude of graduate students in on-campus vs. online statistics courses. *Journal of Statistics Education*, *18*(1), 1-15. https://doi.org/10.1080/10691898.2010.11889472

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) report: A Pre-K-12 curriculum framework*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/gaise/gaiseprek-12_full.pdf

Fullerton, J., & Umphrey, D. (2016). Statistics anxiety and math aversion among advertising students. *Journal of Advertising Education, 20*(1-2), 135-143. https://doi.org/10.1177/10980482160201-216

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22-38. https://doi.org/10.52041/serj.v2i1.557

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372-396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Gibeau, R.-M., & Cousineau, D. (2025). Uncovering key statistical concepts from an investigation of four standardized tests. *Teaching Statistics*, *47*, 219-234.

https://doi.org/10.1111/test.12410

Gibeau, R.-M., Maloney, E. A., Béland, S., Lalande, D., Cantinotti, M., Williot, A., Chanquoy, L., Simon, J., Boislard-Pépin, M.-A., & Cousineau, D. (2023). The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender. *Journal of Numerical Cognition*, *9*(1), 16-43. https://doi.org/10.5964/jnc.8199

Harth, H. (2018). *An analysis of the teaching of introductory statistics at university in "context"* [Doctoral dissertation, Loughborough University]. https://doi.org/10.26174/thesis.lboro.8246996.v1

Husereau, T., Cousineau, D., Zang, S., & Gibeau, R.-M. (2024). A compendium of common heuristics, misconceptions, and biased reasoning used in statistical thinking. *The Quantitative Methods for Psychology*, *20*(1), 57-75. https://doi.org/10.20982/tqmp.20.1.p057

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan. https://us.macmillan.com/books/9780374533557/thinkingfastandslow

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430-454. https://doi.org/10.1016/0010-0285(72)90016-3

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1), 1-9. https://doi.org/10.1080/10691898.1995.11910479

Leder, G. (1985). Sex-related differences in mathematics: An overview. *Educational Studies in Mathematics, 16*, 304-309. https://doi.org/10.1007/BF00776737

Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition*, *4*(3), Article 3.

https://doi.org/10.5964/jnc.v4i3.154

Mackinnon, S. P., Alexander, S. M., Chen, R., Cribbie, R. A., Flett, G. L., & Hill, T. G. (2025). Perfectionism, anxiety sensitivity, and statistics anxiety: A test of the vulnerability-stress model using a 2-wave longitudinal study. *Journal of Research in Personality, 116*, Article 104611. https://doi.org/10.1016/j.jrp.2025.104611

Malone, C. J., Gabrosek, J., Curtiss, P., & Race, M. (2010). Resequencing topics in an introductory applied statistics course. *The American Statistician*, *64*(1), 52-58. https://doi.org/10.1198/tast.2009.08090

Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician*, *64*(2), 97-107. https://doi.org/10.1198/tast.2010.09132

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346-362. https://doi.org/10.1037/0022-0663.93.2.346

Shaughnessy, J. M. (2019). Recommendations about the big ideas in statistics education: A retrospective from curriculum and research. *Cuadernos de Investigación y Formación En Educación Matemática*, *14*(18), Article 18.

Snedecor, G. W. (1948). A proposed basic course in statistics. *Journal of the American Statistical Association*, *43*(241), 53-60. https://doi.org/10.1080/01621459.1948.10483249

Taherdoost, H. (2022). Designing a questionnaire for a research paper: A comprehensive guide to design and develop an effective questionnaire. *Asian Journal of Managerial Science*, *11*(1), 8-16. https://doi.org/10.51983/ajms-2022.11.1.3087

Tucker, A. W., Allendoerfer, C. B., Douglas, E. C., Fehr, H. F., Hildebrandt, M., Meder, A. E., Meister, M., Mosteller, F., Nothrop, E. P., Ranucci, E. R., Rourke, R. E. K., Thomas Jr., G.

B., Van Engen, H., & Wilks, S. S. (1959). *Program for college preparatory mathematics*. College Entrance Examination Board. http://archive.org/details/program-for-college-preparatory-mathematics-1959

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. https://www.jstor.org/stable/1738360

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicothema*, *20*(1), 174-180. https://reunido.uniovi.es/index.php/PST/article/view/8638

Wang, L. (2020). Mediation Relationships Among Gender, Spatial Ability, Math Anxiety, and Math Achievement. *Educational Psychology Review, 32*, 1-15. https://doi.org/10.1007/s10648-019-09487-z

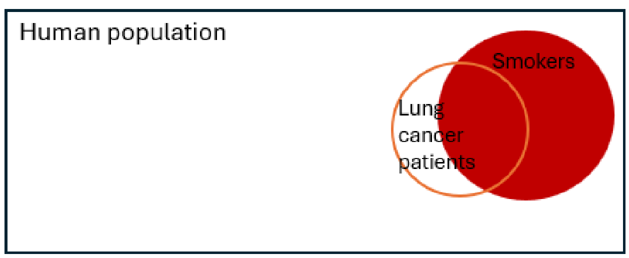
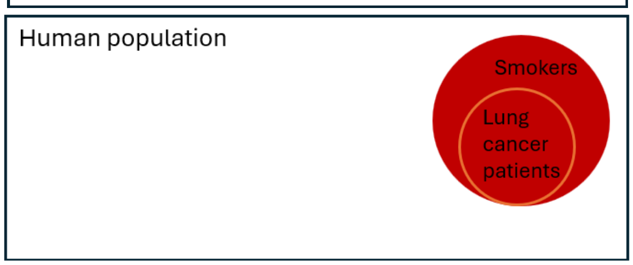
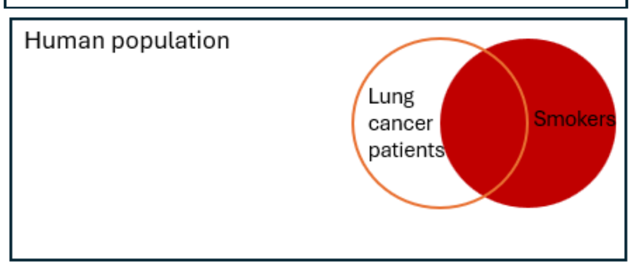
Zavez, K., & Harel, O. (in press). Teaching statistical concepts using computing tools: A review of the literature. *Journal of Statistics and Data Science Education,* 1-12. https://doi.org/10.1080/26939169.2024.2445541

Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 37-70). Springer International Publishing. https://doi.org/10.1007/978-3-319-66195-7_2

Tables

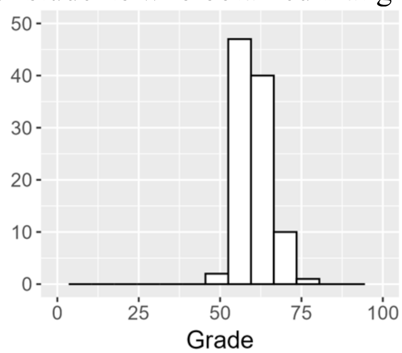
Table 1

The 13 Items of YACUSA Iteration 6

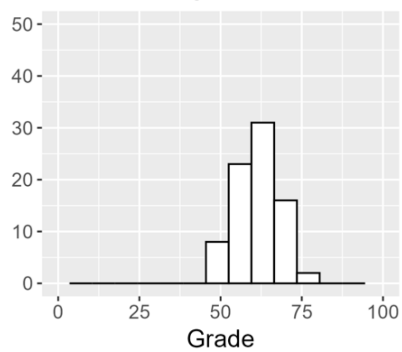
Item Number	Item Presented
1	Let's represent the human population by a box containing all human beings, and all smokers by a red circle, like this:  According to research, the majority of people with lung cancer are smokers. However, the majority of smokers do not necessarily have lung cancer. If we represent all lung cancer patients by an orange circle, what would be the new picture?  a)  b)  c)

d)

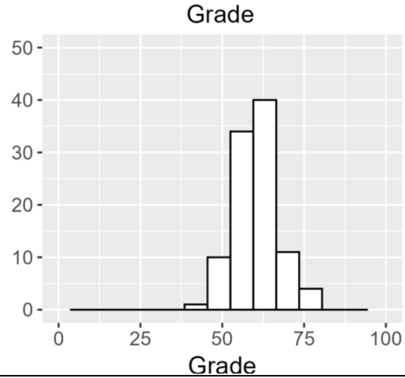
2 In a class, all students have similar grades on the midterm exam. Which of these four frequency graphs show the grades from that class? (A frequency graph shows on the horizontal axis the grades; and on the vertical axis the count of students who obtained that grade.)



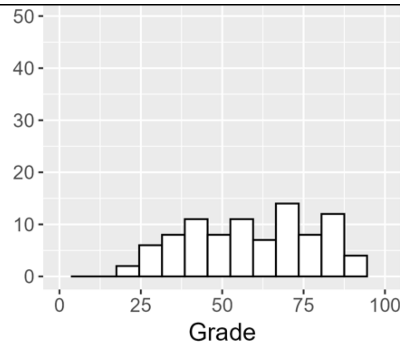
a)



b)

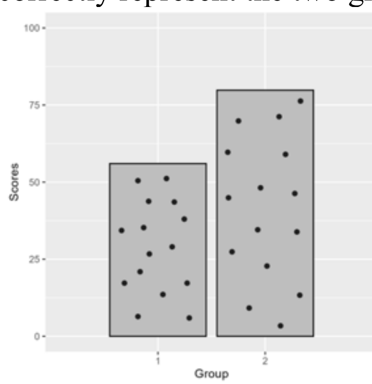


c)

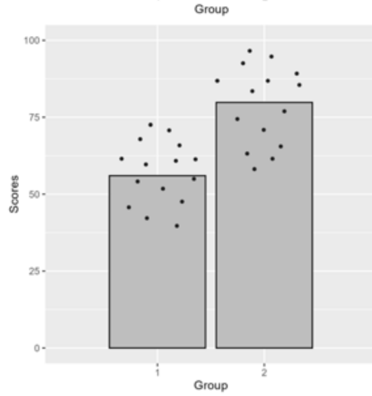


d)

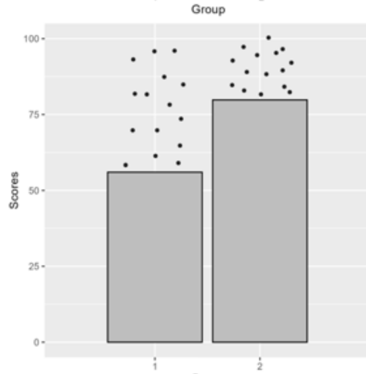
3 The following graphs show the means of scores on a test (vertical axis) for the same two groups (horizontal axis) of people. The mean of Group 1 is 54 and the mean of Group 2 is 77. In each group, there are 15 people. If we were to illustrate the score of each person of each group by a small dot, which graph would correctly represent the two groups?



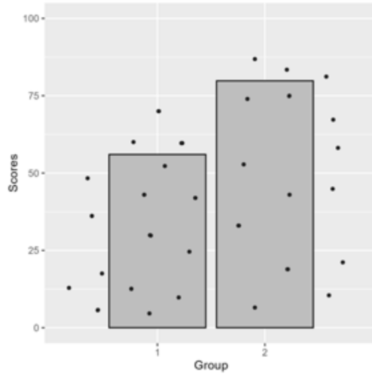
a)



b)

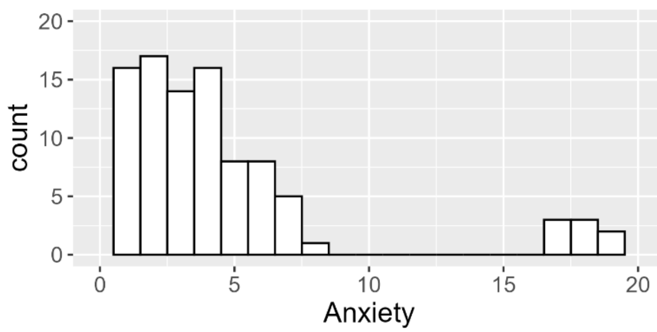


c)



d)

4 An anxiety researcher looks at people's anxiety levels after going to the gym, which should be lower. The researcher sends her research assistants to wait by the door of a gym to stop people on their way out. However, a research assistant misunderstood and waited by the door of an examination classroom. Below is the graph showing people's anxiety levels.



If we ignore people who were stopped after an exam, the average is 3 out of 20. What is the average if we consider all participants?

a) The average would now be approximately 15.
b) The average would still be 3.
c) The average would now be approximately 9
d) The average would now be approximately 5.

5 A group of 10 friends gather and they all flip a fair coin at the same time. The coin of one friend got stuck in their sleeve. The coin of 7 friends, out of the remaining 9, lands on Tail. Below is the result of the coin flips.

Friend 1	T
Friend 2	T
Friend 3	T
Friend 4	T
Friend 5	T
Friend 6	T
Friend 7	T
Friend 8	H
Friend 9	H

Friend 10	?

The 10th friend now flips their coin. On which side will it most likely land on?

a) Head is most likely. Both sides of the coin have a 50% chance, meaning each side should come up once every 2 flips.
b) Tail is most likely. Both sides of the coin should not come up an equal number of times since it is a fair coin.
c) Both sides of the coin are equally likely because each flip has a 50% chance of landing on Head or Tail.
d) Both sides of the fair coin are equally likely because it is impossible to rule out one side of the coin or the other.

6 Which of the following sequences is most likely to result from flipping a fair coin 10 times? Note that H means “Head” and T means “Tail”.

a) H-T-H-H-T-H-H-T-T-H
b) T-H-H-T-T-T-H-H-T-H
c) T-H-T-T-T-T-H-T-T-T
d) All sequences are equally likely because each side has a 1 in 2 chances in each flip.
e) All sequences are equally likely because it is impossible to rule out one or more sequences.

7 Which of the following sequences is least likely to result from flipping a fair coin 10 times? Note that H means “Head and T means “Tail”.

a) H-T-H-H-T-H-H-T-T-H
b) T-H-H-T-T-T-H-H-T-H
c) T-H-T-T-T-T-H-T-T-T
d) All sequences are equally unlikely because each side has a 1 in 2 chances in each flip.
e) All sequences are equally unlikely because it is impossible to rule out one or more sequences.

8 A marketing employee for a streaming platform wants to know if a newly released series is liked or not. To do that, every time the first episode is finished, an appreciation question ranging from 0 (hated it) to 5 (loved it) is asked to the viewer. After looking at the data, the employee concludes that the series is not liked and should be canceled. However, if it is canceled, some viewers will not be happy. Which of the following statements is correct?

a) The viewers who answered the appreciation question all responded between 1 and 4.
b) The viewers who answered the appreciation question mostly responded between 0 and 3.
c) The viewers who answered the appreciation question all responded

between 0 and 2.

d) The viewers who answered the appreciation question mostly responded between 3 and 5.

9 Professor Deschamps needs to send her students' final grade to the administration. The administration asks to divide students into two equal groups –the low achievers (with lower grades) and the high achievers (with higher grades)– and send them the grade that separates the two groups. Facing the table of grades in ascending order below, what is the grade the professors may send to the administration?

55	61	68	70	80	80	86	92

a) The professor may send 75 because it cuts the table in two equal groups of low- and high-achievers.
b) The professor may send 75 because it represents a B+, which is between low and high achievement.
c) The professor may send 80 because it is an A−, which separates achievement levels in two groups.
d) The professor may send 80 because it is the most common grade with two students having this grade.

10 A university often offers grants to students. There are four types of grant given each year: $500 for students showing the most perseverance; $1000 for students who are the most involved; $2000 for students who do the most volunteering; and $4000 for students with the highest GPA. Below is the table of grants given this year.

Number of perseverance grants	15
Number of involvement grants	6
Number of volunteering grants	6
Number of GPA grants	5

Is the average number of grants given the same as the most common number across the different types of grants?

a) Yes, the most common number of grants given is the same as the average number. Both are of 6 grants.
b) No, the most common number of grants given is 6. The average is higher because of the number of perseverance grants.
c) Yes, the average number and the most common number of grants are the same. Both are of 8 grants.

	d) No, the average number of grants given is 6. The most common number is higher because of the overall number of grants given.
11	A finance professor wonders if there is a relation between her students' grades and the amount of money in their saving accounts. She took her students' final grade and asked them how much there was in their savings. The results suggest a strong relation between the students' grades and the amount of money in their savings. What could the results mean? a) Her students' final grades and their savings are perfectly related. b) The higher the grades are, the higher the savings are likely to be. c) An increase in her student's grades increases their savings. d) The analysis is not conclusive.
12	A 1st-year PhD student is interested in the relation between self-confidence and anxiety. While writing the manuscript describing the study, she argues that higher self-confidence leads to lower anxiety. What will her supervisor tell her? a) Looking at the relation between self-confidence and anxiety the way you did does not allow one to draw this kind of conclusion. One can only conclude that they are related. b) Great job! Self -confidence does cause anxiety symptoms to decrease. c) Good job. However, use the term "influences" instead of "leads to". It shows a better understanding of the analysis without compromising your conclusion. d) The results are insufficient to draw any conclusion. We need to repeat the study with a larger number of people who participate.
13	Dr Paulis notices that there is a negative association between the sky during winter and the temperature outside. What does this mean concretely? a) It means that when the sky is clear and blue, the temperature outside gets very cold. b) It does not mean anything other than Dr Paulis's belief. c) It means that the clouds prevent the sun rays from coming in, which decreases the temperature. d) It means that the clearer the sky is, the colder the temperature outside gets or vice versa

Table 2

Factor loadings of the Exploratory Factor Analysis (EFA)

	F 1	F 2	F 3
Item 1	0.38*	0.02	0.16
	[0.14, 0.62]	[−0.09, 0.13]	[−0.10, 0.41]
Item 2	0.33*	−0.09	0.18
	[0.09, 0.57]	[−0.19, 0.02]	[−0.10, 0.46]
Item 3	−0.14	0.04	0.03
	[−0.30, 0.02]	[−0.09, 0.17]	[−0.23, 0.28]
Item 4	−0.52	−0.08	0.10
	[−0.76, −0.28]	[−0.22, 0.06]	[−0.16, 0.36]
Item 5	−0.12	0.10	0.34
	[−0.48, 0.24]	[−0.07, 0.27]	[−0.06, 0.74]
Item 6	0.13	0.44*	0.10
	[−0.05, 0.31]	[0.09, 0.79]	[−0.10, 0.30]
Item 7	−0.01	0.93*	−0.01
	[−0.04, 0.02]	[0.30, 1.56]	[−0.05, 0.03]
Item 8	0.04	−0.01	0.40*
	[−0.29, 0.38]	[−0.12, 0.09]	[0.05, 0.75]
Item 9	0.28	−0.03	−0.26
	[−0.02, 0.57]	[−0.17, 0.12]	[−0.60, 0.09]
Item 10	0.002	−0.02	0.04
	[−0.15, 0.16]	[−0.13, 0.10]	[−0.17, 0.24]

Item 11	0.08	0.001	0.05
	[−0.08, 0.24]	[−0.12, 0.12]	[−0.18, 0.28]
Item 12	0.29*	−0.17*	0.01
	[0.13, 0.45]	[−0.32, -0.02]	[−0.27, 0.30]
Item 13	−0.13	0.04	−0.04
	[−0.28, 0.03]	[−0.09, 0.17]	[−0.26, 0.18]

Table 3

Factor loadings of the Confirmatory Factor Analysis (CFA)

	F 1	F 2	F 3
Item 1	0.38*		
	[0.20, 0.55]		
Item 2	0.33*		
	[0.13, 0.52]		
Item 3	−0.11		
	[−0.26, 0.04]		
Item 4	−0.51*		
	[−0.70, −0.33]		
Item 5		0.07*	
		[0.002, 0.14]	
Item 6		0.56*	
		[0.35, 0.78]	
Item 7		0.72*	
		[0.45, 1.01]	
Item 8			−0.06
			[−0.19, 0.08]
Item 9			0.35*
			[0.15, 0.55]
Item 10			0.04
			[−0.10, 0.19]

Item 11	0.17
	[−0.02, 0.37]
Item 12	0.47*
	[0.22, 0.73]
Item 13	−0.11
	[−0.26, 0.05]

Table 4

Correlation Matrix on YACUSA Score, Last Statistics Grade, Anxiety types and YACUSA's Subscales

Variables	YACUSA	Stat. Grd	SAS-I	SAQ-I	SAQ-M	SAQ-tot	YAC-VR	YAC-RP	YAC-Co
YACUSA	1								
Stat. Grd	0.19	1							
	[0.09, 0.29]								
SAS-I	−0.21	−0.18	1						
	[−0.30, −0.11]	[−0.28, −0.08]							
SAQ-I	−0.04	0.0004	0.49	1					
	[−0.15, 0.06]	[−0.10, 0.10]	[0.40, 0.56]						
SAQ-M	−0.19	−0.07	0.54	0.47	1				
	[−0.29, −0.09]	[−0.17, 0.03]	[0.46, 0.61]	[0.39, 0.55]					
SAQ-tot	−0.15	−0.04	0.60	0.82	0.89	1			
	[−0.24, −0.04]	[−0.15, 0.06]	[0.53, 0.66]	[0.79, 0.85]	[0.87, 0.91]				
YAC-VR	0.63	0.23	−0.22	−0.05	−0.21	−0.16	1		
	[0.56, 0.69]	[0.13, 0.32]	[−0.31, −0.12]	[−0.15, 0.05]	[−0.31, −0.11]	[−0.26, −0.06]			

YAC-PR	0.48	−0.004	−0.09	0.02	−0.03	−0.01	0.01	1	
	[0.39, 0.55]	[−0.11, 0.10]	[−0.19, 0.02]	[−0.09, 0.12]	[−0.14, 0.07]	[−0.12, 0.09]	[−0.09, 0.11]		
YAC-Co	0.76	0.13	−0.10	−0.04	−0.11	−0.09	0.22	0.04	1
	[0.71, 0.80]	[0.02, 0.23]	[−0.20, 0.002]	[−0.14, 0.06]	[−0.21, −0.01]	[−0.19, 0.01]	[0.12, 0.31]	[−0.06, 0.14]	

Notes. YACUSA = Participants' score on the YACUSA; Stat. Grd = Participants' self-reported grade in their last statistics course; SAS-I = Participants' score on the statistics interpretation subscale of the *Statistical Anxiety Scale*; SAQ-I = Participants' score on the imagery subscale of the *Spatial Anxiety Questionnaire*; SAQ-M = Participants' score on the mental manipulation subscale of the *Spatial Anxiety Questionnaire*; SAQ-tot = Participants' computed score on the imagery and mental manipulation subscales of the *Spatial Anxiety Questionnaire*; YAC-VR = The Visual Representation subscale of YACUSA; YAC-PR = The Probability & Randomness subscale of YACUSA; YAC-Co = The Commitment to an Interpretation subscale of YACUSA. The numbers between brackets are the 95% confidence intervals

Table 5

Results of the IRT model

Subscale	Items	Alpha	Beta
Visual	Item 1	1.599	-0.900
Representation	Item 2	0.806	-1.060
	Item 3	0.526	0.689
	Item 4	1.141	0.760
Randomness	Item 5	0.780	-2.233
& Probability	Item 6	7.511	-0.493
	Item 7	3.399	-0.184
Commit	Item 8	0.617	-0.921
	Item 9	0.744	-1.076
	Item 10	0.998	-0.937
	Item 11	0.977	0.237
	Item 12	0.920	0.726
	Item 13	0.632	1.647

Figures

Figure 1

An Example Item from Allen's (2006) on the Development of the Statistics Conceptual Inventory (SCI)

New #7 (Old #11)

Which of the following statistics is least impacted by extreme outliers?

a) Range
b) 3rd quartile
c) Mean
d) Variance

Correct Answer: B
Topic Area: Descriptive

Figure 2

Item Information Curve of the Visual Representation subscale. The item numbers correspond to those of the YACUSA.

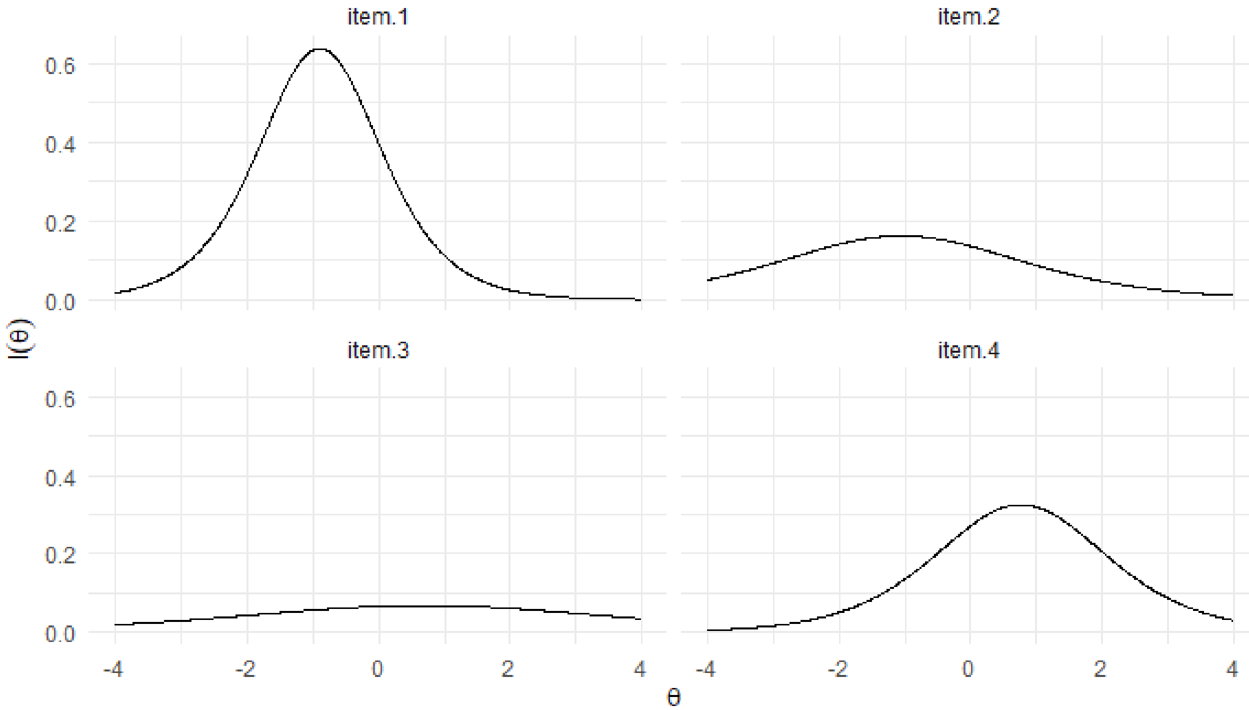


Figure 3

Item Characteristic Curve of the Visual Representation subscale. The item numbers correspond to those of the YACUSA.

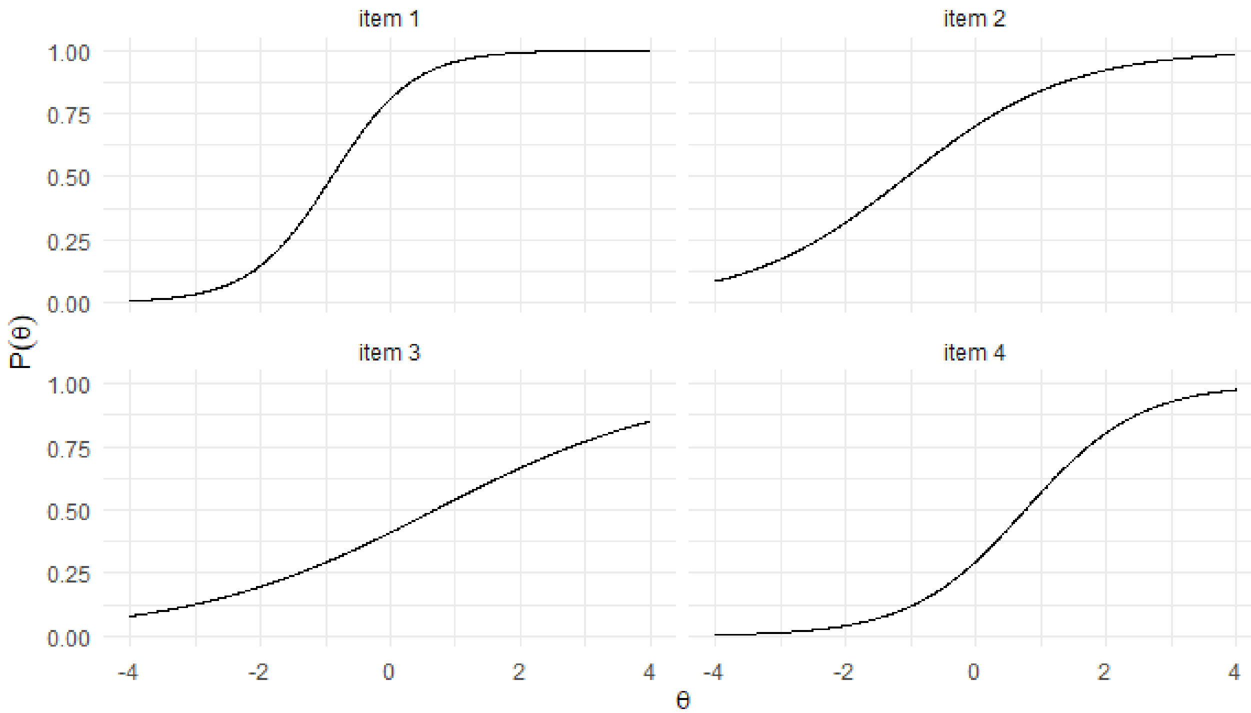


Figure 4

Item Information Curve of the Randomness & Probability subscale. Item 1 = item 5 of the YACUSA, item 2 = item 6, item 3 = item 7

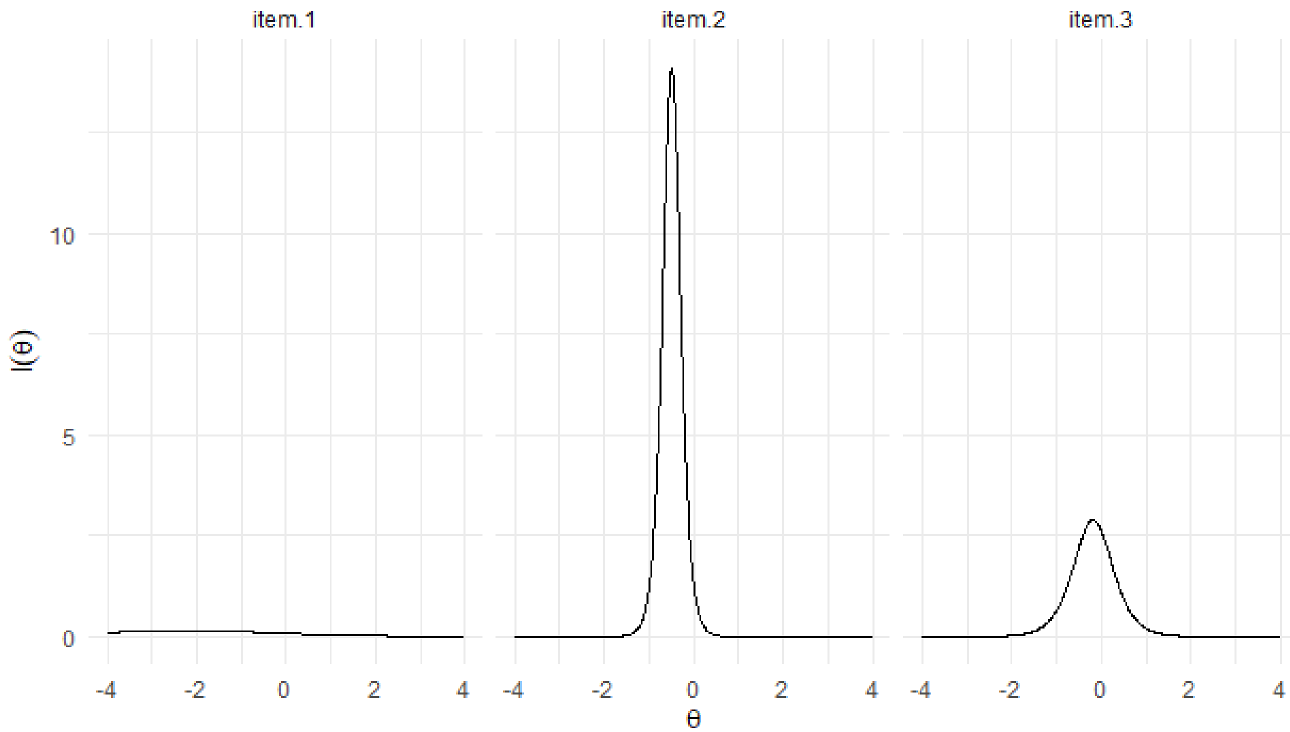


Figure 5

Item Characteristic Curve of the Randomness & Probability subscale. Item 1 = item 5 of the YACUSA, item 2 = item 6, item 3 = item 7

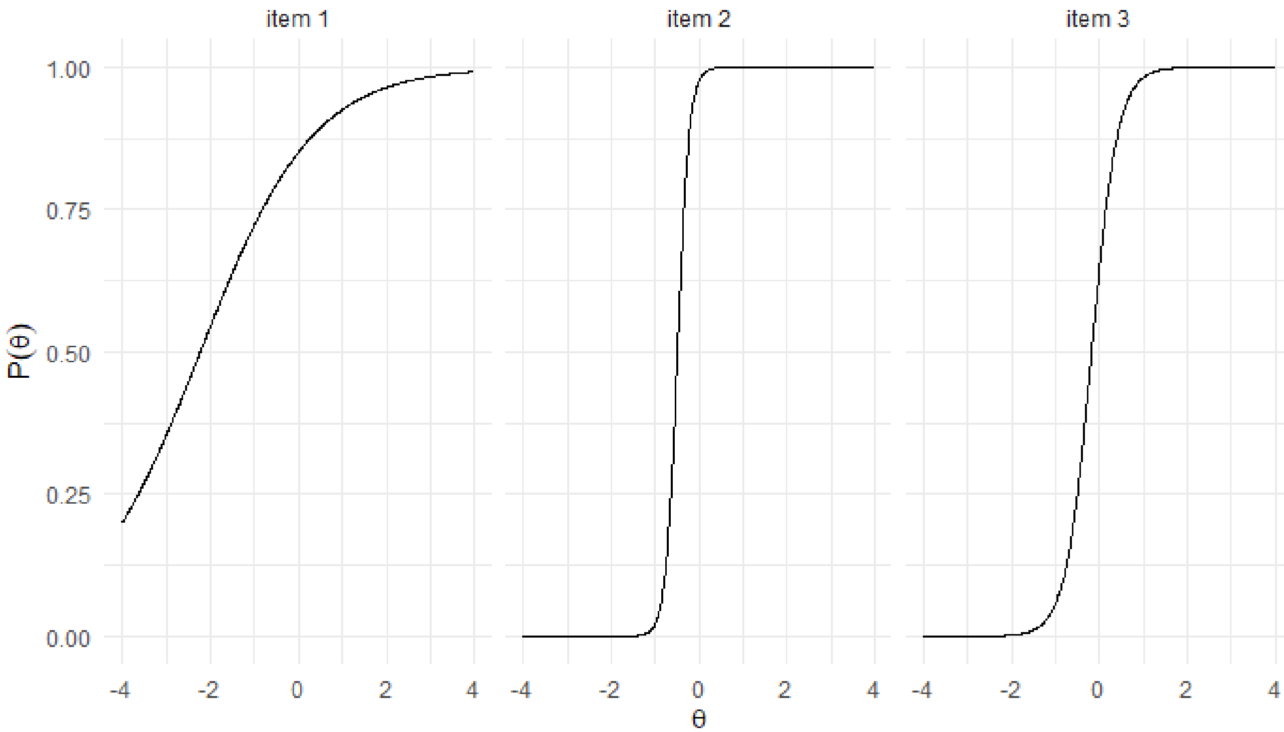


Figure 6

Item Information Curve of the Commit subscale. Item 1 = item 8 of YACUSA, item 2 = item 9, item 3 = item 10, item 4 = item 11, item 5 = item 12, and item 6 = item 13.

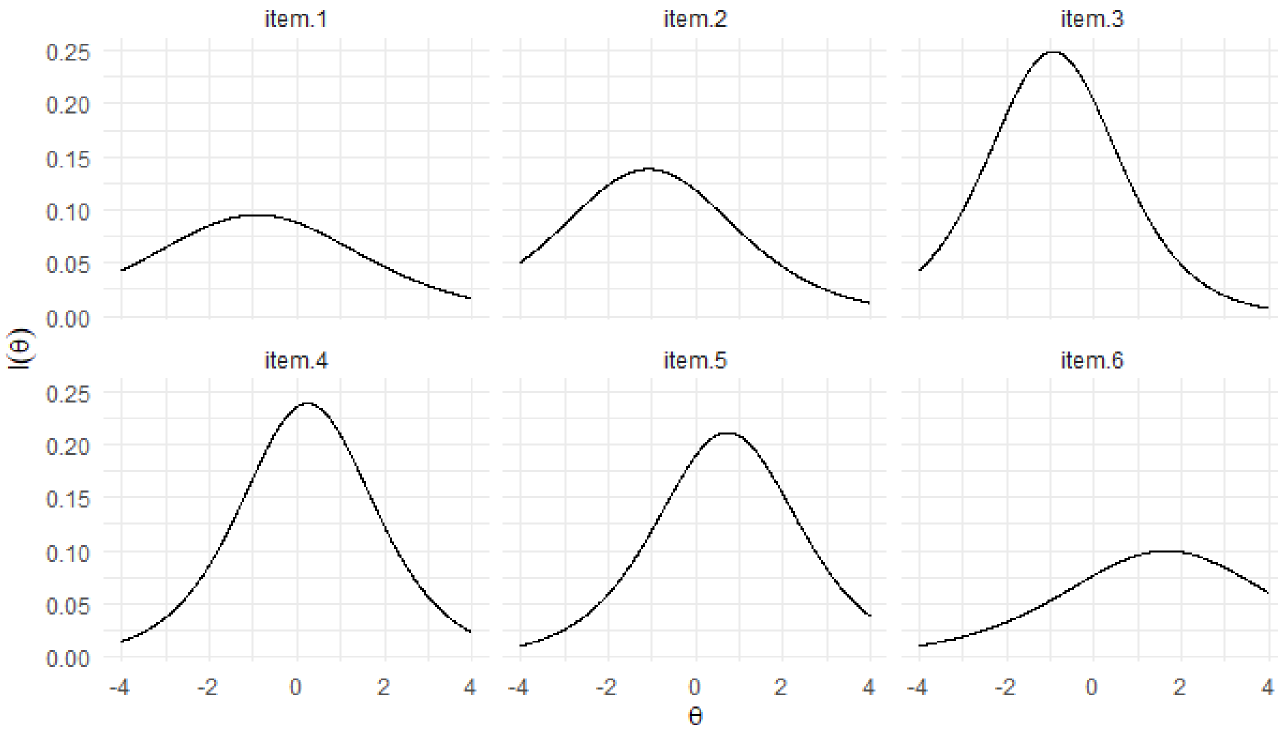
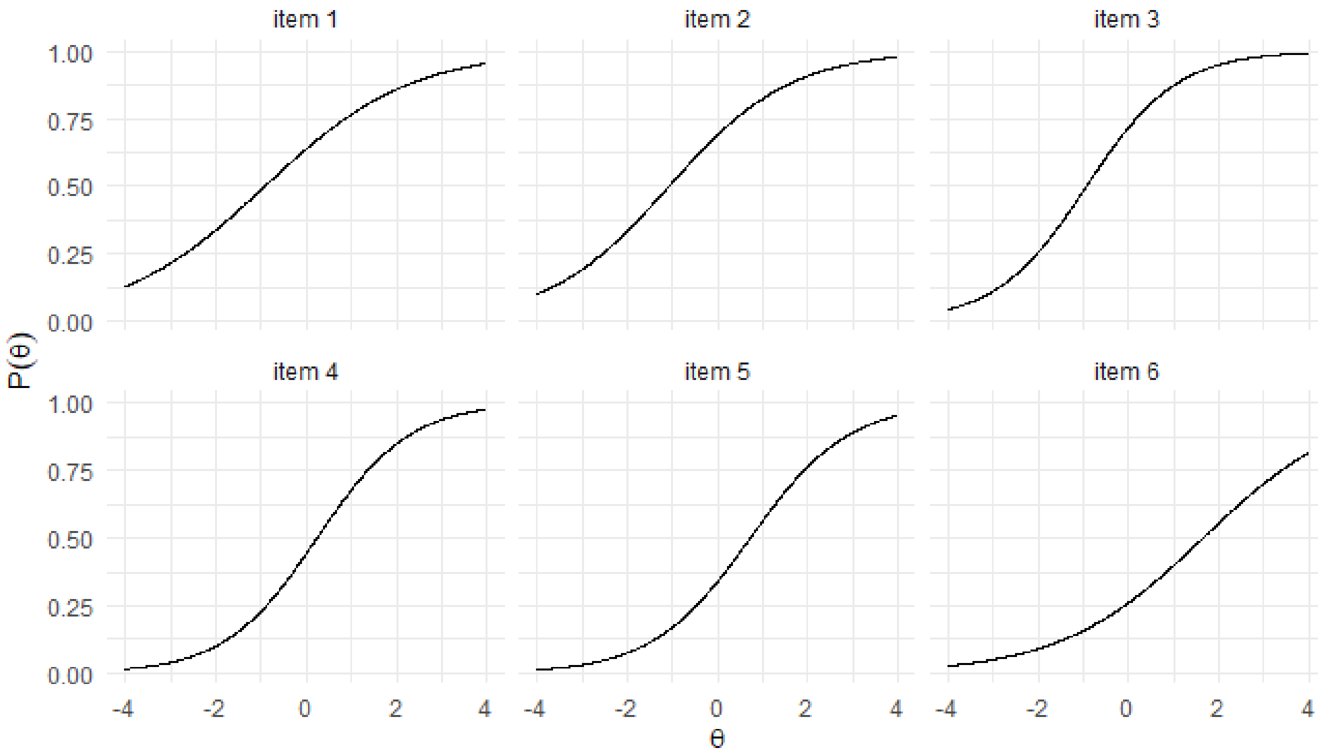


Figure 7

Item Characteristic Curve for the Commit subscale. Item 1 = item 8 of YACUSA, item 2 = item 9, item 3 = item 10, item 4 = item 11, item 5 = item 12, and item 6 = item 13.



Chapter 6: General Discussion

Recapitulating the Thesis Studies and their Contributions

The network of relations and the influence of statistics anxiety on the understanding and performance of statistics is only now emerging despite the research on statistics education dating back to the late nineteenth century (Zieffler et al., 2018). We know many things about this phenomenon, like the contexts in which it manifests, which is not limited to statistics classrooms as statistical information are often presented in the media (Hoffrage et al., 2000). It is also known what it is related to, and what outcome it best predicts. For example, one of the strongest correlates of statistics anxiety is mathematics anxiety (Gibeau et al., 2023; Zeidner, 1991), alongside mathematics performance (Esnard et al., 2021). It is also related to gender (Baloğlu, 2003), general anxiety (Baloğlu, 1999), and is best predicted by performance in statistics courses (Cantinotti et al., 2017; McIntee et al., 2022). However, Mackinnon et al. (2025) reported that performance is a good predictor of statistics anxiety. What is less known, however, is why it is so wildly present and how it came to be. The first two studies of the present thesis tried to answer this.

Studies A and B

It started with a simple thought: if mathematics is related of statistics and to spatial processing, statistics must be related to spatial processing as well. Spatial skills are at the center of most scientific and technological training, abilities and professions (Hinze et al., 2013). It is then logical that it influences statistics training, as it is a tool often used in scientific research. The same can be said about spatial anxiety, as a skill and its anxiety are often sides of the same coin.

Methodology of Study A

The goal of this study was to investigate the relations between spatial anxiety, mathematics anxiety and statistics anxiety. It was hypothesized that mathematics anxiety is a mediator between the other two variables and that spatial anxiety influences the levels of statistics anxiety. To test these hypotheses, we asked 778 university students in social sciences from three different countries if situations about learning and being evaluated in statistics and mathematics and being faced with activities requiring spatial skills make them anxious.

Results of Study A

Correlations revealed that all three anxiety types are strongly associated (rs = .49 and .79; Table 5 of the study). Between the instruments' subscales, the correlations range from .22 to .79 (Table 7). The lowest correlations involve spatial navigation and spatial imagery anxieties. The strongest correlations are between spatial manipulation and statistics interpretation anxieties, as well as mathematics evaluation and statistics interpretation anxieties, and between mathematics learning and statistics evaluation anxieties.

The mediation analyses revealed that two-third of the variance in statistics anxiety can be explained by spatial anxiety both directly and indirectly, through mathematics anxiety, and that more than half of the increase in statistics anxiety can be indirectly explained by spatial anxiety, trough mathematics anxiety (Figure 1 of Study A). Furthermore, multiple regression analyses revealed that the best predictors of statistics interpretation anxiety are mathematics and spatial manipulation anxieties.

Implications of Study A

The results indicate that not only are spatial, mathematics and statistics anxieties strongly correlated, but also that statistics anxiety is predicted by the other two. Specifically, students with higher apprehension toward spatial operations, tend to have higher apprehension toward mathematics and toward statistics. This suggests that if we can lower students' spatial anxiety, their statistics anxiety will also diminish, allowing them to better perform in their course as statistics anxiety is the strongest predictor of statistics performance (Cantinotti et al., 2017; McIntee et al., 2022). Note that this is a causal hypothesis and the results presented in the present thesis cannot support it. Developmental longitudinal studies are needed to test this hypothesis.

Methodology of Study B

This study has two sub-studies. In sub-study 1, the relations between visuospatial working memory span, spatial anxiety, mathematics anxiety and statistics anxiety were examined. The goal was to examine the role of spatial processing skills onto statistics anxiety through spatial and mathematics anxieties. It was hypothesized that visuospatial working memory, compared to verbal working memory, contributes to spatial anxiety, which, in turn, directly and indirectly (through mathematics anxiety) influences statistics anxiety. It was thought that lesser competence in retaining visuospatial information generates anxiety in spatial processing operations, which influences statistics anxiety. To test these hypotheses, we asked 306 university students in social sciences from three different countries if statistical, mathematical and spatial situations make them anxious, and to complete a visuospatial working memory and a verbal working memory task.

In sub-study 2, the goal and the hypotheses were similar as in sub-study 1. Instead of looking at visuospatial working memory, we looked at spatial problem-solving skills. We asked a sample of 374 students of the same population if statistical interpretation, spatial manipulation and spatial imagery situations make them anxious, and to complete a test measuring their spatial and verbal skills.

Results of Study B

In sub-study 1, correlations revealed that all three anxiety types are again strongly associated (*r*s between .49 and .74; Table 2 of the study). Between the anxiety types and the tasks, the correlations range from −.24 to −.10. Only one of these correlations had zero included in the 95% confidence interval, which involves spatial anxiety and verbal working memory. Between the anxiety subscales and the tasks, the correlations range from −.24 to .04 (Table 3). The strongest negative correlations are between visuospatial working memory and mathematics evaluation anxiety, statistics evaluation anxiety, spatial mental manipulation anxiety, mathematics learning anxiety, as well as statistics interpretation anxiety.

The mediation analyses of sub-study 1 revealed that a little less than half of the variance in statistics interpretation anxiety can be explained by spatial mental manipulation anxiety both directly and indirectly, through mathematics anxiety, while considering both working memory spans (Figure 2). However, only visuospatial working memory has a significant prediction onto spatial mental manipulation anxiety.

In sub-study 2, correlations revealed that both anxiety types are moderately associated with spatial skills (*r*s = −.18 for spatial anxiety and −.16 for statistics anxiety; Table 6 of Study B of this thesis). Between the anxiety subscales and spatial skills, the correlations range from −.33 to .05 (Table 7). The strongest negative correlation is between spatial mental manipulation skills and spatial imagery skills with spatial mental manipulation anxiety, spatial imagery anxiety, as well as statistics interpretation anxiety. The mediation analyses revealed that a little less than half of the variance in statistics interpretation anxiety can be explained by spatial mental manipulation skills indirectly, through spatial mental manipulation anxiety (Figure 3).

These results indicate that the capacity to retain visuospatial information as well as spatial skills predict the anxiety toward mentally manipulating elements, which directly and indirectly predicts students' anxiety levels in statistics.

Implications of Study B

The results of sub-study 1 indicate that visuospatial working memory, spatial anxiety, mathematics anxiety and statistics anxiety are related. Furthermore, statistics anxiety is predicted by spatial anxiety, which is predicted by visuospatial working memory. In other words, students with a better capacity to retain visuospatial information, tend to have a lesser apprehension toward spatial operations, and a lesser apprehension toward statistics. This suggests that students with better capacity to retain visuospatial information will perform better in their statistics course.

The results of sub-study 2 indicate that spatial problem solving has a similar role as visuospatial working memory in spatial anxiety and statistics anxiety. Considering that spatial problem solving skills are malleable (Burte et al., 2017), students can be trained in spatial skills. This suggests that training students can lower their spatial anxiety, lower their statistics anxiety, and improve their performance in statistics courses. As above, this is only an hypothesis that needs the be tested with developmental longitudinal studies.

Taken together, these results indicate that spatial processing skills, spatial anxiety, and mathematics anxiety (but not necessarily in that order) influence the amount of statistics anxiety felt by undergraduate students, especially when it comes to interpreting statistical results. This gives an answer key as to why statistics anxiety is present and how it came to be. Indeed, it seems to be influenced by people's competence in and apprehension toward spatial operations and then by their apprehension toward mathematics. To truly test the hypothesis of how statistics anxiety came to be, developmental longitudinal research is needed. However, these results support the pertinence of investigating this avenue.

Understanding its relations with space and mathematics gives at least a chance to statistics instructors to help their students navigate this fear so they can better develop their statistical skills and reasoning. It also indicates that spatial operations, mathematics and statistics have similar underlying cognitive processes. To examine the possibility of common processes, research relating to their respective competence is needed. Study C and D of the present thesis go in that direction.

Study(ies) C (and D)

Studies C and D of the present thesis gave us the opportunity to study the variables' competences (instead of their anxieties; e.g., spatial processing skills instead of spatial anxiety) by investigating existing tests on statistics understanding and by trying to develop a test on conceptual understanding of statistics. Study C gives another answer key as to which concepts such a test should measure. Unfortunately, Study D was not successful. This does not mean that such a test cannot be developed with a better and more robust methodology.

With Study C as the first building block examining four statistical tests development, and research on misconceptions and on the network of relations, an investigation of common cognitive processes and of how statistics anxiety develops in students, additional knowledge on the teaching and learning of statistics can evolve.

Methodology of Study C

The goal of this study was to investigate the statistical concepts measured in tests on statistics understanding from a teachers' perspective. No hypothesis was formed as it is an exploratory study. We assembled the *Statistical Reasoning Assessment* (SRA; Garfield, 2003), the *Statistics Concept Inventory* (SCI; Allen, 2006), the *Comprehensive Assessment of Outcomes for a First Course in Statistics* (CAOS; DelMas et al., 2007), and the *Statistical Reasoning in Biology Concept Inventory* (SRBCI; Deane et al., 2016). We asked 42 instructors who have previously taught at least one statistics course to list the concepts that are being measured in each item of these tests. A thematic analysis was done on their responses.

Results of study C

The concepts

Through the examination of common themes across participants' response on each item, between three and 13 statistical concepts were identified for each test (8 for the SRA, 12 for the SCI, 13 for the CAOS and 3 for the SRBCI). After discussions and revisions, a global list of 18 statistical concepts emerged. These concepts were then defined and categorized either as procedural (i.e., needing step-by-step procedures) or as conceptual (i.e., needing a deep understanding of the signification and implication of the concept). The ones considered as procedural were bolded in Table 3 of the corresponding study.

Participants' answers

During their participation, instructors had the opportunity to answer the items themselves. Basic frequencies and their total score were computed. For each item, many participants, despite their knowledge and expertise in statistics, did not have the correct answer. The test with the highest average score was the SRBCI with 83%. This is unexpected as this test was developed for biology and most participants are psychology instructors. Furthermore, many expressed their disagreement and confusion about the items and their response choices.

Implications of Study C

The results of this study indicate that even tests claiming to measure conceptual understanding of statistics still have items on procedural skills. Also, not all statistically knowledgeable researchers and professors agree with each other on an item's correct answer. This suggests that statistics are not mechanical with only one answer for each problem like mathematics. Indeed, there is humanity in statistics, where different experiences, different knowledge and different perspectives bring different answers or interpretations to the same problem. How we can teach this *humane statistics* to students remains an unexplored question. How we can assess students without procedural skills is the question tackled by Study D of the present thesis.

Methodology of Study D

The goal of this study was to develop a test measuring conceptual understanding of statistics in university students. We thought that by using no formula, no technical terms (e.g., *sample*) and as little numbers as possible, we would grasp the signification and implication of concepts instead of procedural skills. We created items based on previous studies and on experience. A total of six iterations of item testing and revision were run from 2022 to 2025. In each iteration, we asked undergraduate students enrolled in multiple programs to answer the items to the best of their ability.

Results of Study D

The results of this study indicate that the *Yet Another Conceptual Understanding of Statistics Assessment* (YACUSA) is partially valid and unreliable due to a low intra-class correlation and low internal consistencies. Also, most items have a ceiling effect but have good differentiation between all ability levels. The factor structure is not optimal, and more research should look into statistical concepts categorization to improve future test development on conceptual understanding of statistics. The best factor structure found was of three factors that seem to assemble essential statistical skills. Indeed, factor 1 included items with graphs, factor 2 included items on probability, and factor 3 included items where participants needed to choose between possible interpretations or explain why a given interpretation is good or bad. We named and defined those factors as *i)* the understanding of Visual Representations (VR), *ii)* the understanding of Probability, Chance and Randomness (PR), and *iii)* the capacity to Commit to an interpretation (Co).

Implications of Study D

The results of this study indicate that the YACUSA is not appropriate to measure conceptual understanding of statistics. This may be due to an insufficient methodology. Another possibility is that statistics cannot be taught nor assessed by concept. Instead, teaching and assessing by skill might be best. This encourages educators and instructors to rethink and reorganize statistics education.

Concluding Remarks

Based on previous research and on the contributions of the present thesis, it is clear that statistics education is not optimal for today's society. In this STEM (Science, Technology, Engineering and Mathematics) and Big Data era, statistics education needs to be rethought to give students the best training and the best opportunity in understanding the signification and implications of statistical concepts and of the statistical conceptual map (i.e., how the concepts relate to and influence each other). It is no longer sufficient to teach statistics the way our instructors taught us as the society's values and priorities have changed and are continually evolving.

For example, it is no longer recommended to teach with formulas and assess students' capacity to memorize and apply them (e.g., Carver et al., 2016). It is also recommended to add variability, which is often missing in the curriculum (Meletiou, 2000). Many more recommendations have been proposed over the years, yet statistics education slowly changes and no-longer needed procedural skills are still being taught and assessed. This results in less-thanoptimally trained citizens and professionals.

The present thesis suggests an interesting way of teaching statistics at the college or university level: through visuospatial operations and by using a differently organized curriculum. There are a few studies in the last 10 to 15 years that investigated the influence of spatial activities in mathematics learning (e.g., Burte et al., 2017; Cheng & Mix, 2014; Miller & Halpern, 2013). These studies indicate that spatial skills can be improved, which, in turn, improves performance in mathematics or other STEM fields. Burte et al. (2017) implemented the Think3d! program in classes for grades three to six in the United States. This program uses activities like origami and folding, unfolding and cutting paper to develop children's visuospatial thinking. Burte et al. (2017), asked 86 children to complete a spatial skills test and a mathematics test, asked them to participate in Think3d! for six weeks, and retested the children's spatial and mathematics skills. The researchers concluded that the Think3d! program improved children's spatial thinking skills, their mathematics accuracy in problems that involved visuospatial skills only in grades five and six, and their mathematics performance in real-world problems. The results also indicated that improvement in a paper folding task predicts improvement in mathematics.

Most of these studies examined children, most (if not all) looked at mathematics, and most spatial training effect were not long-term (e.g., Miller & Halpern, 2013). However, if both spatial processing and mathematics are related to statistics anxiety, they must also be related to statistics understanding, implying that spatial training may also improve performance in statistics. Using visuospatial operations (e.g., graphs) to teach statistics may improve students' spatial skills and, in turn, improve their understanding and performance as well as reduce their anxiety. Similarly, asking students to do simple spatial tasks (e.g., origami) at the beginning of every class, may improve their spatial skills and, in turn, improve their statistics understanding and performance. To investigate these hypotheses, an instrument measuring conceptual understanding of statistics is needed.

Research on how spatial processing skills and statistics understanding are related is the next step. Furthermore, research is needed on spatial training in adults as most studies were done on children. Indeed, we do not know if spatial skills are malleable in adults and how that can be done. To improve statistics education, long-term longitudinal and experimental research on the influence of spatial processing skill onto statistics learning and understanding is needed.

The present thesis suggests a second way to teach statistics. Indeed, with Study D, it seems that statistics may be best taught as a series of essential statistical skills. For example, instead of teaching distribution, central tendency, variability and confidence intervals separately, it may be best to teach those concepts as part of a skill: visual representation. Similarly, instead of teaching hypothesis testing, errors, and confounding variables, sample/sampling and population separately, it may be best to teach them as part of the capacity to form and commit to an interpretation. A skill-based curriculum may be a good alternative to teach a subject with intertwined concepts.

General Limitations

The studies presented here have a few common limitations. These are summarized in this sub-section. Firstly, three of the four studies have a sample composed mostly of women. There is an extensive amount of literature indicating that anxiety affects women a lot more compared to men (e.g., Arcand et al., 2020). This could have exacerbated the levels of anxiety reported in the present studies as well as strengthening the relations between anxiety types. Indeed, the relations between the anxiety types may solely be based of the gender effect. Relatedly, the majority of the samples herein are composed of social sciences students. Considering that these students may not be science- or math-oriented (compared to STEM fields students), this also may have exacerbated anxiety levels.

Furthermore, some instruments used herein, and the analyses used to examine their validity are limited. For example, the *Spatial Reasoning Instrument* (Ramful et al., 2017) used in Study B may be too easy for the sample as we observed a ceiling effect. To examine the instruments' validity, only one type of validity was used (i.e., construct validity). Other types (e.g., divergent validity) were not examined, which may have influenced the studies' conclusions.

Lastly, many concluding remarks and conjectures were made about causal relations between the variables. However, all the studies presented here are correlational. As such, these remarks and conjectures cannot be taken as this thesis' conclusions or implications. They can be considered as ideas for future research.

Take-Class Message

The present thesis contributes only a small part to the research on statistics education. It brings big ideas through modest, yet robust (for the first three studies) findings, over six years of research. It is the hope that this thesis encourages researchers, educators, and instructors to work together pushing the boundaries to improve and optimize the teaching and learning of statistics. The results of this thesis are a small step but allow future researchers to make bigger steps toward the improvement of statistics education.

Introduction and Discussion References

Adams, P. A., & Adams, J. K. (1958). Training in confidence-judgments. *The American Journal of Psychology*, *71*(4), 747-751. https://doi.org/10.2307/1420334

Allen, K. (2006). *The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics* [Doctoral dissertation, University of Oklahoma]. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2130143

Almarzouki, A. F. (2024). Stress, working memory, and academic performance: A neuroscience perspective. *Stress*, *27*(1), Article 2364333. https://doi.org/10.1080/10253890.2024.2364333

Ashkenazi, S., & Velner, H. (2023). The interplay between math performances, spatial abilities, and affective factors: The role of task. *Trends in Neuroscience and Education*, *33*, Article 100211. https://doi.org/10.1016/j.tine.2023.100211

Baloğlu, M. (1999). A comparison of mathematics anxiety and statistics anxiety in relation to general anxiety. *ERIC*, 1–31. https://eric.ed.gov/?id=ED436703

Baloğlu, M. (2003). Individual differences in statistics anxiety among college students. *Personality and Individual Differences, 34*(5), 855-865. https://doi.org/10.1016/S0191-8869(02)00076-4

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II). A Framework for Statistics and Data Science Education.* American Statistical Association.

Belton, K., & Dhami, M. K. (2020). Cognitive biases and debiasing relevant to intelligence analysis. In R. Viale (Ed.), *Handbook on Bounded Rationality* (pp. 540-553). Routledge.

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *The Journal of Experimental Education*, *57*(3), 247-261. https://doi.org/10.1080/00220973.1989.10806509

Bergman Nutley, S., & Söderqvist, S. (2017). How is working memory training likely to influence academic performance? Current evidence and methodological considerations. *Frontiers in Psychology*, *8*, Article 69. https://doi.org/10.3389/fpsyg.2017.00069

Beyth-Marom, R., Fidler, F. & Cumming, G. (2008). Statistical cognition: Towards evidence-based practice in statistics and statistics education. *Statistics Education Research Journal, 7*(2), 20-39. https://doi.org/10.52041/serj.v7i2.468

Billard, L. (1998). The role of statistics and the statistician. *The American Statistician*, *52*(4), 319-324. https://doi.org/10.1080/00031305.1998.10480589

Boissin, E., Caparos, S., & De Neys, W. (2022). Examining the role of deliberation in de-Bias training. *Social Science Research Network*, 1-41. https://papers.ssrn.com/abstract=4106107

Boissin, E., Caparos, S., Voudouri, A., & De Neys, W. (2022). Debiasing System 1: Training favours logical over stereotypical intuiting. *Judgment and Decision Making, 17*(4), 646-690. http://dx.doi.org/10.2139/ssrn.4106107

Broers, N. J. (2006, July 2-7). *Learning goals: The primacy of statistical knowledge* [conference presentation]. In 7th International Conference on Teaching Statistics (ICOTS 7), Salvador, Brazil.

Burte, H., Gardony, A. L., Hutton, A., & Taylor, H. A. (2017). Think3d!: Improving mathematics learning through embodied spatial training. *Cognitive Research: Principles and Implications*, *2*, Article 13. https://doi.org/10.1186/s41235-017-0052-9

Cantinotti, M., Lalande, D., Ferlatte, M.-A., & Cousineau, D. (2017). Validation de la version francophone du questionnaire d'anxiété statistique (SAS-F-24). *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *49*(2), 133-142. https://doi.org/10.1037/cbs0000074

Caro-Barrera, J. R., García-Moreno García, M. B., Jara Alba, C. A., & Pérez-Priego, M. (2023, July 3-5). *The importance of data visualisation in introductory statistics courses in social sciences degrees* [conference presentation]. In 15th International Conference on Education and New Learning Technologies, Palma, Spain. https://doi.org/10.21125/edulearn.2023.1303

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Roswell, G. H., Velleman, P., Witmer, J., & Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report 2016*. American Statistical Association. https://commons.erau.edu/publication/1083

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2-11. https://doi.org/10.1080/15248372.2012.725186

Coulter, X. (1995). A new way of teaching statistics. *Journal of Educational Technology Systems*, *24*(2), 181-187. https://doi.org/10.2190/K906-0M2X-CGTE-7XGV

Cousineau, D. (2020). In-class activity comparing standard errors as a function of sample size with SPSS. *The Quantitative Methods for Psychology, 16*(2), v4-v7. https://doi.org/10.20982/tqmp.16.2.v009

Cousineau, D., & Harding, B. (2017). Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions. *Revue de Psychoéducation, 46*(2), 397-419. https://doi.org/10.7202/1042257ar

Cui, X., & Guo, K. (2022). Supporting mathematics learning: A review of spatial abilities from research to practice. *Current Opinion in Behavioral Sciences, 46*, Article 101176. https://doi.org/10.1016/j.cobeha.2022.101176

Curry, L. (2002). *Learning statistics at a distance* [master's dissertation, Massey University]. http://hdl.handle.net/10179/10945

de Hevia, M. D., Vallar, G., & Girelli, L. (2008). Visualizing numbers in the mind's eye: The role of visuo-spatial processes in numerical abilities. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1361-1372. https://doi.org/10.1016/j.neubiorev.2008.05.015

Deane, T., Nomme, K., Jeffery, E., Pollock, C., & Birol, G. (2016). Development of the Statistical Reasoning in Biology Concept Inventory (SRBCI). *CBE—Life Sciences Education*, *15*(1), 1-13. https://doi.org/10.1187/cbe.15-06-0131

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371-396. https://doi.org/10.1037/0096-3445.122.3.371

Delage, V., Trudel, G., Retanal, F., & Maloney, E. A. (2022). Spatial anxiety and spatial ability: Mediators of gender differences in math anxiety. *Journal of Experimental Psychology: General*, *151*(4), 921-933. http://dx.doi.org/10.1037/xge0000884

DelMas, R. C. (2002). Statistical literacy, reasoning, and thinking: A commentary. *Journal of Statistics Education*, *10*(2), 1-11. https://doi.org/10.1080/10691898.2002.11910674

DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28-58. http://www.stat.auckland.ac.nz/serj

Dunn, P. K. (2005). We can still learn about probability by rolling dice and tossing coins.

Teaching Statistics, *27*(2), 37-41. https://doi.org/10.1111/j.1467-9639.2005.00205.x

Edgeworth, F. Y. (1885). Methods of statistics. *Journal of the Statistical Society of London*, 181-217.

Esnard, T. R., Alladin, F. M., & Samlal, K. H. (2021). Prior mathematics performance, statistics anxiety, self-efficacy and expectations for performance in statistics: A survey of social sciences students in a Caribbean institution of higher education. *Statistics Education Research Journal, 20*(1), Article 4. https://doi.org/10.52041/serj.v20i1.98

Falk, R., & Konold, C. (1993). Random means hard to digest. *Focus on Learning Problems in Mathematics, 16*(1), 2-12. https://eric.ed.gov/?id=EJ485501

Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences*, *39*, 1-12. https://doi.org/10.1016/j.lindif.2015.02.007

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, *28*(2), 161-186. https://doi.org/10.1016/S0361-476X(02)00015-2

Forbes, S. D. (2012). Data visualisation: A motivational and teaching tool in official statistics. *Technology Innovations in Statistics Education*, *6*(1), 1-19. https://doi.org/10.5070/T561012851

Gal, I., & Garfield, J. B. (1997). *The assessment challenge in statistics education*. IOS Press.

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22-38. https://doi.org/10.52041/serj.v2i1.557

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 32-44. https://doi.org/10.1037/0096-1523.32.1.32

Gibeau, R.-M., Maloney, E. A., Béland, S., Lalande, D., Cantinotti, M., Williot, A., Chanquoy, L., Simon, J., Boislard-Pépin, M.-A., & Cousineau, D. (2023). The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender. *Journal of Numerical Cognition*, *9*(1), 16-43. https://doi.org/10.5964/jnc.8199

Gunderson, E. A., Ramirez, G., Beilock, S., & Levine, S. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, *48*(5), 1229-1241. https://doi.org/10.1037/a0027433

Harth, H. (2018). *An analysis of the teaching of introductory statistics at university in "context"* [Doctoral dissertation, Loughborough University]. https://doi.org/10.26174/thesis.lboro.8246996.v1

Hayslett, H. T. (1968). *Statistics made simple* (P. Murphy, Ed.; 3rd ed.). Elsevier.

Hinze, S. R., Williamson, V. M., Shultz, M. J., Williamson, K. C., Deslongchamps, G., & Rapp, D. N. (2013). When do spatial abilities support student comprehension of STEM visualizations? *Cognitive Processing, 14*(2), 129-142. https://doi.org/10.1007/s10339-013-0539-3

Holmes, P. J. (2014). *The effect of motivation and attitude towards statistics on conceptual understanding of statistics* [Doctoral dissertation, University of Georgia].

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290*(5500), 2261-2262. https://www.science.org/doi/10.1126/science.290.5500.2261

Hubbard, R. (1997). Assessment and the process of learning statistics. *Journal of Statistics Education*, *5*(1), 1-9. https://doi.org/10.1080/10691898.1997.11910522

Husereau, T., Cousineau, D., Zang, S., & Gibeau, R.-M. (2024). A compendium of common heuristics, misconceptions, and biased reasoning used in statistical thinking. *The Quantitative Methods for Psychology*, *20*(1), 57-75. https://doi.org/10.20982/tqmp.20.1.p057

Jung, S., & Hwang, J. (2016). Students' understanding of statistical terms having lexical ambiguity. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics: International perspectives* (pp. 151-161). Springer International Publishing. https://doi.org/10.1007/978-3-319-23470-0_20

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430-454. https://doi.org/10.1016/0010-0285(72)90016-3

Khoshnoodifar, M., Ashouri, A., & Taheri, M. (2023). Effectiveness of gamification in enhancing learning and attitudes: A study of statistics education for health school students. *Journal of Advances in Medical Education & Professionalism*, *11*(4), 230-239. https://doi.org/10.30476/jamp.2023.98953.1817

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*(1), 59-98. https://doi.org/10.1207/s1532690xci0601_3

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1), 1-9. https://doi.org/10.1080/10691898.1995.11910479

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, *24*(5), 392-414. https://doi.org/10.2307/749150

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology. General*, *139*(4), 665-682. https://doi.org/10.1037/a0020198

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty* (1st ed., pp. 306-334). Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.023

Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition*, *4*(3), Article 3. https://doi.org/10.5964/jnc.v4i3.154

Mackinnon, S. P., Alexander, S. M., Chen, R., Cribbie, R. A., Flett, G. L., & Hill, T. G. (2025). Perfectionism, anxiety sensitivity, and statistics anxiety: A test of the vulnerability-stress model using a 2-wave longitudinal study. *Journal of Research in Personality, 116*, Article 104611. https://doi.org/10.1016/j.jrp.2025.104611

Malone, C. J., Gabrosek, J., Curtiss, P., & Race, M. (2010). Resequencing topics in an introductory applied statistics course. *The American Statistician*, *64*(1), 52-58. https://doi.org/10.1198/tast.2009.08090

Marson, S. M. (2007). Three empirical strategies for teaching statistics. *Journal of Teaching in Social Work*, *27*(3-4), 199-213. https://doi.org/10.1300/J067v27n03_13

McIntee, S.-E., Goulet-Pelletier, J.-C., Williot, A., Deck-Léger, E., Lalande, D., Cantinotti, M., & Cousineau, D. (2022). (Mal)adaptive cognitions as predictors of statistics anxiety. *Statistics Education Research Journal*, *21*(1), Article 5.

https://doi.org/10.52041/serj.v21i1.364

Meletiou, M. M. (2000). *Developing students' conceptions of variation: An untapped well in statistical reasoning* [Doctoral dissertation, University of Texas].

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences*, *26*, 141-152. https://doi.org/10.1016/j.lindif.2012.03.012

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, *65*(2), 123-137. https://doi.org/10.1111/j.1751-5823.1997.tb00390.x

Moore, D. S., & Cobb, G. W. (2000). Statistics and mathematics: Tension and cooperation. *The American Mathematical Monthly*, *107*(7), 615-630. https://doi.org/10.1080/00029890.2000.12005247

Mustafa, R. Y. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistics Education*, *4*(1), 1-10. https://doi.org/10.1080/10691898.1996.11910504

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, *238*(4827), 625-631.

Onwuegbuzie, A. J., & Seaman, M. A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *The Journal of Experimental Education*, *63*(2), 115-124. https://doi.org/10.1080/00220973.1995.9943816

Ouyang, H., Wang, S., & Huang, L. (2025). Survey on the current status of statistical cognition and teaching needs of Chinese medical students. *Frontiers in Public Health, 13*, Article 1621667. https://doi.org/10.3389/fpubh.2025.1621667

Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: students' understanding of the mean. *Educational Studies in Mathematics*, *12*(2), 191-204.

Rice, M. S., Tomlin, G., & Stein, F. (2024). *Clinical research in occupational therapy* (Sixth ed.). Taylor & Francis.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*(1), 33-51. https://doi.org/10.1002/(SICI)1099-0771(199703)10:1<33::AID-BDM244>3.0.CO;2-6

Sokolowski, H. M., Hawes, Z., & Lyons, I. M. (2019). What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition*, *182*, 193-212. https://doi.org/10.1016/j.cognition.2018.10.005

Stuart, M. (1995). Changing the teaching of statistics. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *44*(1), 45-54. https://doi.org/10.2307/2348615

Tomkins, L., & Ulus, E. (2016). ‘Oh, was that “experiential learning”?!’ Spaces, synergies and surprises with Kolb’s learning cycle. *Management Learning*, *47*(2), 158-178. https://doi.org/10.1177/1350507615587451

Toomarian, E. Y., & Hubbard, E. M. (2018). On the genesis of spatial-numerical associations: Evolutionary and cultural factors co-construct the mental number line. *Neuroscience & Biobehavioral Reviews, 90*, 184-199. https://doi.org/10.1016/j.neubiorev.2018.04.010

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110. https://doi.org/10.1037/h0031322

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. https://www.jstor.org/stable/1738360

Wardrop, R. L. (1995). *Statistics: Learning in the presence of variation*. Wm. C. Brown.

Willcox, W. F. (1936). Definitions of statistics. *Revue de l’Institut International de Statistique / Review of the International Statistical Institute*, *3*(4), 388-399.

https://doi.org/10.2307/1401217

Zavez, K., & Harel, O. (in press). Teaching statistical concepts using computing tools: A review of the literature. *Journal of Statistics and Data Science Education,* 1-12. https://doi.org/10.1080/26939169.2024.2445541

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology, 61*(3), 319-328. https://doi.org/10.1111/j.2044-8279.1991.tb00989.x

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, *16*(2), 2-25. https://doi.org/10.1080/10691898.2008.11889566

Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 37-70). Springer International Publishing. https://doi.org/10.1007/978-3-319-66195-7_2

Appendices

Appendix A: Study A Chapter 2 of this Thesis

Appendix A: Translation of the Items for the Three In-House Translations

Table A1

Translation of the Spatial Anxiety Questionnaire (SAQ)

	Original Items	Translated Items
1	Asked to imagine the 3-dimensional structure of a complex molecule using only a 2-dimensional picture for reference	Vous devez imaginer la structure en trois dimensions d'une molécule complexe à partir d'une image en seulement deux dimensions
2	Asked to determine how a series of pulleys will interact given only a 2-dimensional diagram	Vous devez déterminer comment des leviers et des poulies vont interagir à partir d'un diagramme en deux dimensions
3	Asked to imagine and mentally rotate a 3-dimensional figure	Vous devez imaginer et tourner mentalement un objet tridimensionnel
4	Asked to imagine a 3-dimensional structure of the human brain from a 2-dimensional image	Vous devez imaginer la structure en trois dimensions d'un cerveau humain à partir d'une image en deux dimensions
5	Asked to imagine the motion of a mechanical system given a static picture of the system	Vous devez imaginer le mouvement d'un système mécanique à partir d'une image statique de ce système
6	Imagining on a test what a 3-dimensional landscape model would look like from a different point of view	Dans un test, vous devez imaginer à quoi ressemblerait un paysage en trois dimensions si vous le regardiez à partir d'un autre point de vue
7	Asked to imagine the 3-dimensional shape created by rotating a complex 2-dimensional plane on an exam	Dans un examen, vous devez imaginer la forme tridimensionnelle qui est créée en effectuant la rotation d'un plan bidimensionnel complexe
8	Using a 3-dimensional model of an airport to complete a homework assignment	Vous devez utiliser un modèle en trois dimensions d'un aéroport pour terminer un devoir
9	Finding your way to an appointment in an area of a city or town with which you are not familiar.	Vous devez trouver votre chemin pour aller à un rendez-vous dans une partie d'une ville que vous ne connaissez pas sans utiliser de carte ni de smartphone
10	Finding your way back to your hotel after becoming lost in a new city.	Vous devez retrouver votre chemin vers votre hôtel après vous être perdu(e) dans une ville que vous ne connaissez pas sans utiliser de carte ni

		de smartphone
11	Asked to follow directions to a location across town without the use of a map	Vous devez suivre des indications pour atteindre un lieu de l'autre côté de la ville sans utiliser de carte ni de smartphone
12	Finding your way back to a familiar area after realizing you have made a wrong turn and become lost while driving.	Après avoir réalisé que vous avez pris un mauvais tournant et vous être perdu(e) en conduisant, vous devez retrouver votre chemin vers un lieu familier sans utiliser de carte ni de smartphone ni de GPS
13	Trying to get somewhere you have never been to before in the middle of an unfamiliar city.	Vous devez vous rendre à un endroit où vous n'êtes jamais allé(e) auparavant dans un une ville qui ne vous est pas familière sans utiliser de carte ni de smartphone
14	Trying a new route that you think will be a shortcut without the benefit of a map.	Vous devez essayer un nouvel itinéraire qui serait plus court selon vous sans l'aide d'une carte ni de smartphone
15	Asked to do the navigational planning for a long car trip	Vous devez planifier l'itinéraire pour un long voyage en auto sans utiliser de carte ni de smartphone ni de GPS
16	Memorizing routes and landmarks on a map for an upcoming exam	Vous devez mémoriser les routes et les repères sur une carte routière pour un examen à venir
17	Asked to recall the shade and pattern of a person's tie you met for the first time the previous evening.	Vous devez vous rappeler la couleur et les motifs de la cravate d'une personne que vous avez rencontrée pour la première fois la veille
18	Asked to give a detailed description of a person's face whom you've only met once.	Vous devez donner une description détaillée du visage d'une personne que vous avez rencontré une seule fois
19	Asked to recall the exact details of a relative's face whom you have not seen in several years.	Vous devez donner des détails précis sur le visage d'un membre de votre famille que vous n'avez pas vu depuis des années
20	Asked to recreate your favorite artist's signature from memory	Vous devez imiter de mémoire la signature de votre artiste préféré
21	Describing in detail the cover of a book to a bookseller because you've forgotten both the title and author of the book.	Vous devez décrire en détails la couverture d'un livre à un libraire parce que vous avez oublié le titre et l'auteur de l'ouvrage
22	Tested on your ability to create a drawing or painting that reproduces the details of	Vous êtes testé(e) sur votre habileté à produire une peinture ou un dessin qui reproduit les

	a photograph as precisely as possible.	détails d'une photographie aussi précisément que possible
23	Asked to imagine and describe the appearance of a radio announcer or someone you've never actually seen.	Vous devez imaginer et décrire l'apparence d'un animateur de radio ou de quelqu'un que vous n'avez jamais vu
24	Given a test in which you were allowed to look at and memorize a picture for a few minutes, and then given a new, similar picture and asked to point out any differences between the two pictures	Lors d'un test, vous devez regarder et mémoriser une image pendant quelques minutes et ensuite, visualiser une seconde image semblable et identifier les différences entre les deux images

Note. For the Belgian questionnaire, the items were ungendered

Table A2

Translation of the Abbreviated Mathematics Anxiety Scale (AMAS)

	Original Items	Translated Items
1	Having to use the tables in the back of a math book	Vous devez utiliser les tables se trouvant à la fin d'un livre de mathématiques
2	Thinking about an upcoming math test 1 day before	Vous pensez à un test de mathématiques un jour avant celui-ci.
3	Watching a teacher work an algebraic equation on the blackboard	Vous regardez un enseignant résoudre une équation au tableau.
4	Taking an examination in a math course	Vous faites un examen dans un cours de mathématiques
5	Being given a homework assignment of many difficult problems that is due the next class meeting	Vous devez faire un devoir de mathématiques contenant plusieurs problèmes difficiles qui est à remettre lors du prochain cours.
6	Listening to a lecture in math class	Vous écoutez une leçon de mathématiques en classe.
7	Listening to another student explain a math formula	Vous écoutez un autre étudiant expliquer une formule mathématique
8	Being given a “pop” quiz in math class	Vous recevez un quiz surprise dans un cours de mathématiques
9	Starting a new chapter in a math book	Vous débutez la lecture d'un nouveau chapitre dans un livre de mathématiques

Note. For the Belgian questionnaire, the items were ungendered.

Table A3

Translation of the Single-Item Mathematics Anxiety (SIMA) Scale

	Original Item	Translated Item
1	On a scale from 1 to 10, how math anxious are you?	Sur une échelle de 1 à 10, à quel point êtes-vous anxieux(se) des mathématiques?

Note. For the Belgian questionnaire, the items were ungendered.

Appendix B: Study D Chapter 3 of this Thesis

Iteration 1

Method

Participants. Students enrolled in an introductory to psychology course were recruited via the ISPR system of the author's affiliation. It was expected to recruit 500 participants. This number was based on a previous study (i.e., Lyons et al., 2018) that recruited 485 participants. Students had until the end of classes of the Winter semester of 2022 to complete the study. A total of 310 participants were included in the analyses.

Sociodemographic questions were asked to participants of this first version. Among these participants, 64.3% identified as women, 31% identified as men, 1.2% identified as non-binary, 0.9% identified as gender fluid and the rest of the participants selected "other" (0.2%) or "prefer not to say" (2.1%). Most participants (93%) are in the age group 18-24 years old, 3% are between 25-34 years old, 2.6% are between 35-44 years old, 1.2% are between 45-54 years old and 0.2% are between 55-64 years old. Participants are from a variety of programs. They could select their area of study from 33 programs offered at the authors' affiliation. Most participants are studying biomedical science (21.4%) and psychology (21.7%), and some are studying health sciences (8.4%) and criminology (5.8%). The other participants are distributed between natural sciences, social sciences and other health or medical programs (1 to 3.3%).

Materials. A total of 53 multiple-choice items were created and divided into 16 subscales (see Table 1 of the main text for a list of concepts and their definition). The concept Po has 4 items (items 1 to 4), RP has 6 items (items 5 to 10), DRC has 4 items (items 11 to 14), RRR has 2 items (items 15 and 16), Dis has 3 items (items 17 to 19), S has 4 items (items 20 to 23), Co has 4 items (items 24 to 27), Dr has 2 items (items 28 and 29), A has 3 items (items 30 to 32), VD has 2 items (items 33 and 34), EM has 2 items (items 35 and 36), C has 5 items (items 37 to 41), Ct has 3 items (items 42 to 44), SI has 3 items (items 45 to 47), NP has 5 items (items 48 to 52), M has 1 item (item 53). The correct response options are coded as “1” and the other response options are coded as “0”. This applies to all iterations.

Results

Data Screening. A total of 577 students clicked on the link, 429 completed the questionnaire, and 310 were kept for analysis. Removed participants completed the questionnaire in 9 minutes or less, which is well below the estimation time of 25 minutes. These 310 participants generated 16 106 responses and 324 missing responses. The Little’s test was run to examine if missing values are missing at random. The test was significant (p = .008), meaning that the values are not missing randomly. These missing responses are considered as incorrect and were replaced with “0”. Among the total number of responses, 11 861 were incorrect.

Internal Consistency. To test the questionnaire’s reliability, intra-class correlation (ICC) and the Cronbach’s alpha are used. The McDonald’s omega could not be computed due to a lack of covariance. The ICC is .77 with the 95% CI being [.73, .81], and the alpha is .79. The results indicate that internal consistency of this version is good.

Factor Analyses. To examine the questionnaires’ structure, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were computed. These results are presented in Tables B1.1 to B2. Table B1.1 shows the EFA’s factor loadings of all items on factors 1 to 8. Table B1.2 shows the EFA’s factor loadings of all items on factors 9 to 16. The results indicate that this 16-factor model fits the data very well (RMSEA = .000, CFI = 1.000, TLI = 1.027, SRMR = .041). However, most factor loadings are not significant (i.e., 0 is included in their confidence intervals) and are not in line with the theoretical background. Then, a CFA based on the theoretical background was run. Table B2 shows that some items load strongly on their expected factor while others load weakly or not at all. The fit indices could not be computed because the model lacks identification.

IRT Models. A total of 16 IRT models were run; one per concept. However, the analysis generated no results for *Modeling* (the 16th concept) as it has only one item. Table B3 shows the alpha and beta parameters separated by concept. As can be seen, the slopes vary between −0.23 and 10.77. A little more than half of the items differentiate well between all ability levels (low positive alphas). The difficulties vary between −4.00 and 18.41. A little less than half of the items have a negative beta parameter, meaning they are too easy. Visual representations of these results are shown in Figures B1 to B15.

Iteration 2

Method

Participants. The same population as in iteration 1 was recruited. However, considering that data collection for this iteration was in the summer semester of 2022 and less students are enrolled during this time, a smaller sample size was expected. In ISPR, we set the number of registrations to 500, but only 141 students registered. A total of 91 participants were included in the analyses.

Materials. Five subscales were kept for this iteration: Po (items 1to 4), RP (items 5 to 8), S (items 9 to 11), VD (items 12 to 14), and Ct (items 15 to 17). The first two subscales have four items each and the last three subscales have three items each for a total of 17 items. These five subscales were chosen as it is believed that they are the building blocks of statistics understanding. Indeed, to understand any statistic, one needs to understand that there is a *population* from which we draw a *sample* using one of many (*sampling*) methods. Then, one needs to understand that the sample and its individual responses can be anything within the range of possible values as they are *random* but that their similarities and differences can be visualized and estimated (*variability*, *dispersion* and *central tendencies*) with a degree of certainty (*probability*).

Results

Data Screening. A total of 141 students clicked on the link, 96 completed the questionnaire, and 91 were kept for analysis. Removed participants either did not complete the questionnaire or completed it in 3 minutes or less, which is well below the estimation time of 10 minutes. These 91 participants generated 1 540 responses and 7 missing responses. The Little's test was run to examine if missing values are missing at random. The test was not significant (p = .961). These missing values were replaced by "0". Among all responses, 560 were incorrect.

Internal Consistency. To test the questionnaire's reliability, intra-class correlation (ICC) and the Cronbach's alpha are used. The ICC is .36 with the 95% CI being [.53, .18], and the alpha is .40. The McDonald's omega could not be computed due to a lack of covariance. The results indicate that internal consistency is weak.

Factor Analyses. The results of the factor analyses are presented in Tables B4 and B5. The EFA revealed that the 4-factor model fits the data well (RMSEA = .01, CFI = .99, TLI = .99, SRMR = .08), which was unexpected as there are 5 concepts being measured. A CFA based on the theoretical background of 5 concepts was run. Table B5 shows that some items load strongly on their expected factor while others load weakly or not at all. The fit indices could not be computed because the model lacks identification.

IRT Models. Table B6 shows the alpha and beta parameters separated by concept. As can be seen, the slopes vary between −0.64 to 15.42. About half of the items differentiate well between all ability levels (low positive alpha). The difficulties vary between −3.51 to 2.65. Most items have a negative beta parameter, meaning that they are too easy. Visual representations of these results are shown in Figures B16 to B19.

Iteration 3

Method

Participants. The same population as in iteration 1 was recruited. The expected sample size was the same as previous iterations (i.e., 500 participants) and 414 students clicked on the link during the Winter semester of 2024. A total of 236 were kept for analysis.

Materials. First, the concepts *Population* and *Sample & Sampling* are assembled to form one concept (*Population & Sample* [PoS]). We made this decision based on two reasons; *i*) as researchers, one cannot separate the sample from its population and, most of the time, one cannot separate the population from its sample, and *ii*) analyses of the previous iteration indicate that the four-factor model best fitted the data. Regarding the items, PoS has 13 items (items 1 to 7 and items 17 to 22), RP has 9 items (items 8 to 16), VD has 6 items (items 23 to 28), and Ct has 6 items (items 29 to 34).

Results

Data Screening. A total of 414 students clicked on the link to the questionnaire, 309 completed the questionnaire, and 236 were kept for analysis. Removed participants completed the questionnaire in 9 minutes or less, which is well below the estimated time of 20 minutes. These participants generated 8 024 responses and 98 missing responses. The Little’s test was run to examine if missing values are missing at random. The test was not significant (p = .804). These missing values were replaced with “0”. Among all responses, 2 672 are incorrect.

Internal Consistency. To test the questionnaire’s reliability, intra-class correlation (ICC), the Cronbach’s alpha and the McDonald’s omega are used. The ICC is .75 with the 95% CI being [.79, .70], the alpha is .78, and the omega is .83. The results indicate that internal consistency is good.

Factor Analyses. The results of the factor analyses are presented in Tables B7 and B8. The EFA revealed that the 4-factor model fits the data well (RMSEA = .014, CFI = .990, TLI = .987, SRMR = .083). However, most factor loadings are not significant (i.e., 0 is included in their confidence intervals) and are not in line with the theoretical background. A CFA based on the theoretical background was run. Table B8 shows that the model has an acceptable fit to the data (RMSEA = .041, CFI = .893, TLI = .885, WRMR = 1.134). As can be seen, factor loadings are acceptable at best and negative at worst.

IRT Models. Table B9 shows the alpha and beta parameters separated by concept. As can be seen, the slopes vary between 0.07 and 17.20. A little more than half of the items differentiate well between all ability levels (low positive alpha). The difficulties vary between −2.94 and 7.98. About 3/4 of the items have a negative beta parameter, meaning that they are too easy. Visual representations of these results are shown in Figures B20 to B23.

Iteration 4

Method

Participants. The same population as in iteration 1 was recruited. As in Iteration 2, data collection was in the summer, and a smaller sample size was expected. In ISPR, the number of registration was set at 500, but only 203 students registered to participate. A total of 161 participants were included in the analyses.

Materials. A total of 33 items were presented to participants. PoS has 12 items (items 1to 12), RP has 9 items (items 13 to 21), VD has 6 items (items 22 to 27), and Ct has 6 items (items 28 to 33).

Results

Data screening. A total of 203 students clicked on the link to the test, 182 completed the questionnaire, and 161 were kept for analysis. Removed participants completed the questionnaire in 9 minutes or less, which is well below the estimated time of 20 minutes. Participants generated 5 292 responses and 21 missing values. The Little's test was run to examine if missing values are missing at random. The test was not significant (p = .971). These missing values were replaced with "0". Among all responses, 1 818 were incorrect.

Internal Consistency. To test the questionnaire's reliability, intra-class correlation (ICC) and the Cronbach's alpha are used. The McDonald's omega could not be computed due to a lack of covariance. The ICC is .71 with the 95% CI being [.65, .77], and the alpha is .76. The results indicate that internal consistency is good.

Factor Analyses. The results of the factor analyses are presented in Table B10 and B11. The EFA revealed that the 4-factor model best fits the data (RMSEA = .017, CFI = .968, TLI = .958, SRMR = .115). As can be seen in Table B10, many items do not load significantly on any factor. A CFA was computed based on the theoretical background. Table B11 shows the results. The 4-factor model has an acceptable fit with the data (RMSEA = .034, CFI = .871, TLI = .860, WRMR = 1.015). However, when examining the factor loadings, most are acceptable at best and negative at worst.

IRT Models. Table B12 shows the alpha and beta parameters separated by concept. As can be seen, the slopes vary between −0.31 and 15.29. About half of the items differentiate well between ability levels (low positive alpha). The difficulties vary between −4.66 and 5.22. A little more than 3/4 of the items have a negative beta parameter, meaning that they are too easy. Visual representations of these results are shown in Figures B24 to B27. These results further indicate that the fourth version of YACUSA is not adequate.

Iteration 5

Method

Participants. The same population as in iteration 1 was recruited. The expected sample size was 500, and 458 students clicked on the link during the Winter semester of 2025. A total of 350 participants were included in the analysis.

Materials. One concept was added to this version: *Shared Variance & Correlation* (Co). This concept was added because it is believed that to understand more advanced concepts, one needs to understand that responses can be related to a degree of certainty. A total of 26 items were presented to participants. PoS has 9 items (items 1 to 9), RP has 4 items (items 10 to 13), and VD has 5 items (items 14 to 18), Ct has 4 items (items 19to 22), and Co has 4 items (items 23 to 26).

Results

Data screening. A total of 458 students clicked on the link to the questionnaire, 391 completed it, and 350 were kept for analysis. Removed participants completed the questionnaire in 9 minutes or less, which is well below the estimated time of 19 minutes. Participants generated 9 082 responses and 18 missing values. The Little's test was run to examine if missing values are missing at random. The test was not significant (p = .107). These missing values were replaced with “0”. Among all responses, 3 512 were incorrect responses.

Internal Consistency. To test the questionnaire’s reliability, intra-class correlation (ICC) and the Cronbach’s alpha are used. The McDonald’s omega could not be computed due to a lack of covariance. The ICC is .68 with the 95% CI being [.63, .73], and the alpha is .70. The results indicate that internal consistency is acceptable-to-good.

Factor Analyses. The results of the factor analyses are presented in Table B13 and B15. The EFA revealed that the 3- and 4-factor models are similar. They both fit very well with the data (3-factor model: RMSEA = .021, CFI = .970, TLI = .960, SRMR = .072; 4-factor model: RMSEA = .004, CFI = .999, TLI = .999, SRMR = .061), which is unexpected as there are 5 concepts being measured. A CFA based on the theoretical background (with 5 factors) was run. Table B15 shows that this model has a weak-to-acceptable fit with the data (RMSEA = .049, CFI = .812, TLI = .789, WRMR = 1.248).

IRT Models. Table B16 shows the alpha and beta parameters separated by concept. As can be seen, the slopes vary between −0.35and 10.53. Most items differentiate well between ability levels (low positive alpha), except items 12, 13, 15 and 22. The difficulties vary between −2.97 and 4.36. Most items have a negative beta parameter, except for items 4, 9, 16, 18, 25 and 26. Visual representations of these results are shown in Figures B28 to B32.

Tables of the Appendix B

Table B1.1

Factor Loadings of the Exploratory Factor Analysis of Iteration 1 from Factor 1 to Factor 8

Items	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8
1	−0.15	0.18*	−0.06	−0.10	−0.15	−0.02	−0.02	0.11
	[−0.34, 0.04]	[0.01, 0.35]	[−0.21, 0.09]	[−0.25, 0.05]	[−0.34, 0.04]	[−0.23, 0.20]	[−0.30, 0.27]	[−0.20, 0.42]
2	0.19	0.01	−0.11	−0.01	0.26*	0.10	0.15	0.30*
	[−0.01, 0.39]	[−0.15, 0.17]	[−0.26, 0.03]	[−0.17, 0.15]	[0.05, 0.47]	[−0.09, 0.29]	[−0.07, 0.36]	[0.02, 0.58]
3	0.79*	0.09	0.001	0.01	−0.03	−0.07	0.01	−0.01
	[0.49, 1.09]	[−0.03, 0.20]	[−0.10, 0.10]	[−0.09, 0.10]	[−0.13, 0.07]	[−0.20, 0.05]	[−0.11, 0.14]	[−0.11, 0.09]
4	0.18	−0.19*	0.04	−0.12	0.24	0.23	−0.04	−0.19
	[−0.07, 0.43]	[−0.36, −0.01]	[−0.10, 0.18]	[−0.28, 0.04]	[−0.02, 0.49]	[−0.10, 0.55]	[−0.29, 0.21]	[−0.48, 0.09]
5	0.05	0.84*	0.05	−0.07	0.03	−0.05	−0.11	0.13
	[−0.07, 0.17]	[0.98, 0.69]	[−0.06, 0.15]	[−0.19, 0.03]	[−0.10, 0.16]	[−0.16, 0.06]	[−0.23, 0.02]	[−0.02, 0.28]
6	0.03	0.91*	0.02	0.04	0.11*	0.03	−0.02	−0.12*

	[−0.07, 0.13]	[0.73, 1.08]	[−0.08, 0.11]	[−0.06, 0.13]	[0.001, 0.21]	[−0.07, 0.12]	[−0.13, 0.09]	[−0.22, −0.02]
7	> 0.000	−0.02	0.04	0.98*	0.01	0.04	−0.01	0.02
	[−0.05, 0.05]	[−0.08, 0.04]	[−0.03, 0.11]	[0.76, 1.21]	[−0.04, 0.07]	[−0.04, 0.12]	[−0.07, 0.05]	[−0.04, 0.08]
8	−0.11	0.35*	−0.07	0.06	−0.15	0.03	0.36	−0.14
	[−0.35, 0.12]	[0.16, 0.55]	[−0.23, 0.09]	[−0.10, 0.23]	[−0.34, 0.04]	[−0.35, 0.41]	[−0.05, 0.76]	[−0.39, 0.11]
9	−0.08	0.38*	0.08	0.06	−0.01	0.25	0.18	0.06
	[−0.31, 0.16]	[0.17, 0.58]	[−0.08, 0.23]	[−0.13, 0.25]	[−0.20, 0.19]	[−0.10, 0.60]	[−0.07, 0.43]	[−0.17, 0.29]
10	−0.10	0.02	0.02	0.14	0.07	0.63*	0.04	0.09
	[−0.28, 0.08]	[−0.14, 0.17]	[−0.12, 0.15]	[−0.02, 0.30]	[−0.11, 0.25]	[0.26, 1.00]	[−0.14, 0.21]	[−0.09, 0.27]
11	0.01	0.04	1.20*	0.03	0.04	0.02	−0.002	−0.02
	[−0.05, 0.06]	[−0.02, 0.10]	[0.99, 1.40]	[−0.02, 0.08]	[−0.03, 0.11]	[−0.04, 0.08]	[−0.06, 0.06]	[−0.08, 0.04]
12	0.03	0.13*	−0.04	0.04	0.86*	0.08	0.02	0.09
	[−0.08, 0.13]	[0.02, 0.23]	[−0.16, 0.08]	[−0.04, 0.13]	[0.65, 1.06]	[−0.07, 0.23]	[−0.10, 0.15]	[−0.05, 0.23]
13	−0.05	0.13	0.10	−0.01	0.55*	−0.30	0.14	0.08
	[−0.24, 0.15]	[−0.02, 0.28]	[−0.15, 0.23]	[−0.15, 0.12]	[0.30, 0.80]	[−0.60, 0.004]	[−0.08, 0.35]	[−0.12, 0.27]
14	0.10	0.08	0.06	0.04	−0.04	0.03	0.01	0.10

	[−0.05, 0.25]	[−0.06, 0.21]	[−0.06, 0.17]	[−0.07, 0.14]	[−0.16, 0.07]	[−0.10, 0.15]	[−0.14, 0.17]	[−0.12, 0.33]
15	0.01	−0.04	0.09	−0.11	0.22*	−0.01	0.26*	0.01
	[−0.18, 0.20]	[−0.20, 0.12]	[−0.07, 0.24]	[−0.27, 0.06]	[0.02, 0.42]	[−0.23, 0.20]	[0.05, 0.47]	[−0.21, 0.22]
16	−0.04	−0.15	0.21*	−0.25*	0.06	0.16	0.10	0.20
	[−0.24, 0.17]	[−0.33, 0.03]	[0.07, 0.36]	[−0.44, −0.07]	[−0.14, 0.26]	[−0.14, 0.46]	[−0.14, 0.35]	[−0.27, 0.43]
17	0.04	−0.03	0.01	0.05	0.17	0.05	−0.06	0.72*
	[−0.10, 0.17]	[−0.16, 0.10]	[−0.11, 0.13]	[−0.08, 0.17]	[−0.03, 0.36]	[−0.11, 0.21]	[−0.20, 0.08]	[0.37, 1.07]
18	0.02	0.07	0.04	−0.11	−0.02	−0.17	−0.01	0.31*
	[−0.15, 0.19]	[−0.09, 0.22]	[−0.10, 0.18]	[−0.26, 0.04]	[−0.21, 0.16]	[−0.36, 0.03]	[−0.23, 0.20]	[0.02, 0.60]
19	−0.13	0.13	−0.08	0.13	0.04	−0.03	0.30*	0.28*
	[−0.33, 0.07]	[−0.03, 0.29]	[−0.24, 0.08]	[−0.03, 0.29]	[−0.17, 0.24]	[−0.29, 0.23]	[0.05, 0.56]	[0.05, 0.50]
20	0.02	−0.14*	0.003	0.001	0.14	0.01	0.75*	0.04
	[−0.12, 0.17]	[−0.28, −0.003]	[−0.11, 0.12]	[−0.12, 0.12]	[−0.02, 0.29]	[−0.14, 0.17]	[0.53, 0.97]	[−0.17, 0.25]
21	0.004	−0.01	−0.003	0.02	−0.03	0.01	0.01	−0.02
	[−0.02, 0.03]	[−0.04, 0.03]	[−0.04, 0.03]	[−0.02, 0.05]	[−0.08, 0.02]	[−0.02, 0.03]	[−0.03, 0.05]	[−0.05, 0.02]
22	−0.14	−0.12	0.03	−0.05	0.29*	0.01	−0.01	−0.10

	[−0.35, 0.07]	[−0.29, 0.06]	[−0.14, 0.19]	[−0.21, 0.12]	[0.06, 0.51]	[−0.28, 0.30]	[−0.25, 0.22]	[−0.34, 0.14]
23	−0.02	0.18	0.02	−0.17*	0.07	0.18	0.20	0.24
	[−0.22, 0.18]	[−0.002, 0.37]	[−0.14, 0.18]	[−0.33, −0.01]	[−0.13, 0.28]	[−0.05, 0.40]	[−0.08, 0.48]	[−0.02, 0.50]
24	−0.22*	0.16	0.13	0.05	-0.05	0.10	0.16	0.15
	[−0.43, −0.01]	[−0.02, 0.33]	[−0.03, 0.28]	[−0.11, 0.22]	[−0.25, 0.14]	[−0.12, 0.33]	[−0.08, 0.39]	[−0.12, 0.41]
25	0.05	−0.06	0.10	0.01	−0.05	0.14	0.33*	−0.01
	[−0.16, 0.27]	[−0.22, 0.11]	[−0.08, 0.28]	[−0.15, 0.17]	[−0.25, 0.15]	[−0.08, 0.37]	[0.07, 0.59]	[−0.22, 0.20]
26	0.01	0.16	−0.02	0.13	0.26*	0.22	−0.08	−0.20
	[−0.19, 0.21]	[−0.02, 0.34]	[−0.17, 0.14]	[−0.04, 0.30]	[0.06, 0.47]	[−0.06, 0.49]	[−0.32, 0.15]	[−0.47, 0.07]
27	0.25*	0.15	−0.12	0.02	0.08	0.12	0.36*	0.01
	[0.03, 0.46]	[−0.02, 0.31]	[−0.28, 0.04]	[−0.14, 0.18]	[−0.12, 0.27]	[−0.12, 0.35]	[0.08, 0.65]	[−0.23, 0.25]
28	−0.30*	0.12	−0.03	−0.03	0.13	−0.13	0.24	0.02
	[−0.57, −0.03]	[−0.04, 0.28]	[−0.18, 0.12]	[−0.19, 0.13]	[−0.10, 0.35]	[−0.46, 0.21]	[−0.04, 0.51]	[−0.25, 0.29]
29	−0.01	−0.01	0.01	0.13	−0.21*	0.03	−0.17	0.08
	[−0.21, 0.18]	[−0.18, 0.16]	[−0.14, 0.16]	[−0.03, 0.29]	[−0.40, −0.01]	[−0.19, 0.24]	[−0.42, 0.07]	[−0.15, 0.32]
30	0.07	0.09	0.08	−0.04	−0.03	−0.13	0.24*	0.12

	[−0.12, 0.27]	[−0.05, 0.22]	[−0.04, 0.20]	[−0.17, 0.09]	[−0.20, 0.15]	[−0.42, 0.16]	[0.01, 0.46]	[−0.07, 0.31]
31	−0.01	0.06	0.03	0.11	0.09	0.26	−0.07	−0.24
	[−0.23, 0.21]	[−0.11, 0.23]	[−0.13, 0.18]	[−0.06, 0.29]	[−0.12, 0.30]	[−0.09, 0.60]	[−0.32, 0.19]	[−0.55, 0.08]
32	−0.07	0.05	0.02	−0.13*	0.12	−0.09	0.04	−0.02
	[−0.20, 0.05]	[−0.07, 0.16]	[−0.10, 0.14,]	[−0.25, −0.01]	[−0.03, 0.26]	[−0.24, 0.06]	[−0.09, 0.17]	[−0.13, 0.09]
33	0.07	0.15	0.01	0.04	−0.24*	0.18	0.23	0.14
	[−0.14,0.28]	[−0.04, 0.34]	[−0.18, 0.20]	[−0.14, 0.22]	[−0.47, −0.02]	[−0.09, 0.45]	[−0.08, 0.54]	[−0.12, 0.39]
34	−0.13	0.22*	0.14	−0.02	−0.13	0.02	0.05	0.17
	[−0.33, 0.08]	[0.04, 0.40]	[−0.02, 0.30]	[−0.18, 0.15]	[−0.33, 0.07]	[−0.21, 0.24]	[−0.20, 0.31]	[−0.12, 0.46]
35	0.05	−0.07	0.06	−0.01	−0.24*	0.05	−0.08	−0.06
	[−0.14, 0.24]	[−0.26, 0.12]	[−0.12, 0.23]	[−0.19, 0.16]	[−0.44, −0.03]	[−0.20, 0.30]	[−0.32, 0.16]	[−0.30, 0.19]
36	0.27*	−0.09	0.20*	−0.003	0.28*	0.12	0.18	0.11
	[0.02, 0.53]	[−0.25, 0.07]	[0.02, 0.37]	[−0.15, 0.15]	[0.07, 0.49]	[−0.14, 0.37]	[−0.13, 0.48]	[−0.13, 0.36]
37	−0.09	0.07	−0.06	0.08	−0.07	−0.07	−0.02	0.02
	[−0.21, 0.03]	[−0.04, 0.17]	[−0.17, 0.04]	[−0.01, 0.18]	[−0.15, 0.02]	[−0.20, 0.06]	[−0.11, 0.07]	[−0.07, 0.12]
38	0.11	0.18	0.10	0.10	0.07	−0.30*	0.15	0.09

	[−0.11, 0.32]	[−0.01, 0.37]	[−0.05, 0.26]	[−0.08, 0.27]	[−0.13, 0.28]	[−0.53, −0.06]	[−0.10, 0.39]	[−0.14, 0.32]
39	0.08	−0.01	−0.05	−0.04	0.09	−0.10	0.46*	−0.12
	[−0.14, 0.30]	[−0.18, 0.16]	[−0.19, 0.09]	[−0.20, 0.12]	[−0.11, 0.30]	[−0.35, 0.16]	[0.21, 0.72]	[−0.35, 0.11]
40	0.14	0.13	−0.15	−0.09	−0.04	0.23	0.40*	0.03
	[−0.09, 0.37]	[−0.04, 0.31]	[−0.31, 0.01]	[−0.26, 0.08]	[−0.26, 0.18]	[−0.05, 0.50]	[0.06, 0.73]	[−0.22, 0.28]
41	0.02	−0.08	0.07	0.15	−0.04	0.18	0.26	0.09
	[−0.18, 0.21]	[−0.24, 0.08]	[−0.08, 0.23]	[−0.01, 0.30]	[−0.22, 0.15]	[−0.06, 0.42]	[−0.02, 0.54]	[−0.14, 0.31]
42	0.04	−0.03	−0.06	−0.02	−0.02	0.03	0.08	−0.22
	[−0.17, 0.26]	[−0.20, 0.13]	[−0.21, 0.09]	[−0.18, 0.13]	[−0.21, 0.18]	[−0.21, 0.27]	[−0.15, 0.30]	[−0.51, 0.06]
43	−0.11	0.03	−0.02	−0.07	−0.09	0.05	−0.13	0.14
	[−0.32, 0.11]	[−0.16, 0.21]	[−0.18, 0.13]	[−0.24, 0.09]	[−0.29, 0.12]	[−0.18, 0.29]	[−0.35, 0.10]	[−0.11, 0.40]
44	−0.05	−0.02	−0.16	0.28*	0.11	−0.30	0.10	0.02
	[−0.27, 0.16]	[−0.18, 0.14]	[−0.30, −0.01]	[0.09, 0.46]	[−0.09, 0.31]	[−0.67, 0.08]	[−0.16, 0.35]	[−0.19, 0.23]
45	0.06	0.01	0.19	0.25*	0.14	−0.16	0.04	−0.07
	[−0.15, 0.26]	[−0.17, 0.18]	[0.02, 0.35]	[0.06, 0.43]	[−0.06, 0.35]	[−0.37, 0.05]	[−0.17, 0.25]	[−0.29, 0.15]
46	0.01	0.03	0.19*	0.05	0.26*	0.05	0.39*	0.29*

	[−0.17, 0.20]	[−0.13, 0.19]	[0.03, 0.34]	[−0.11, 0.20]	[0.06, 0.46]	[−0.17, 0.27]	[0.09, 0.68]	[0.04, 0.54]
47	0.04	−0.01	0.01	0.02	0.06	0.02	0.36*	0.41*
	[−0.15, 0.22]	[−0.16, 0.14]	[−0.13, 0.14]	[−0.13, 0.17]	[−0.13, 0.25]	[−0.18, 0.23]	[0.05, 0.67]	[0.15, 0.68]
48	−0.08	−0.03	0.42*	−0.04	0.14	−0.08	0.06	−0.02
	[−0.26, 0.09]	[−0.18, 0.11]	[0.19, 0.64]	[−0.16, 0.09]	[−0.05, 0.33]	[−0.27, 0.10]	[−0.15, 0.26]	[−0.20, 0.16]
49	0.10	−0.07	0.03	0.01	−0.06	0.01	−0.16	0.14
	[−0.07, 0.27]	[−0.22, 0.07]	[−0.11, 0.18]	[−0.13, 0.15]	[−0.23, 0.11]	[−0.17, 0.19]	[−0.39, 0.07]	[−0.11, 0.38]
50	−0.13	−0.02	−0.06	0.07	−0.04	−0.02	0.07	0.06
	[−0.34, 0.06]	[−0.18, 0.14]	[−0.23, 0.11]	[−0.10, 0.24]	[−0.23, 0.15]	[−0.29, 0.24]	[−0.16, 0.30]	[−0.19, 0.31]
51	0.25*	−0.03	0.06	0.31*	−0.001	−0.14	0.04	0.03
	[0.04, 0.45]	[−0.20, 0.15]	[−0.08, 0.21]	[0.12, 0.50]	[−0.19, 0.18]	[−0.35, 0.07]	[−0.19, 0.28]	[−0.22, 0.28]
52	−0.001	−0.05	0.81*	−0.01	−0.17*	−0.05	−0.03	0.06
	[−0.15, 0.15]	[−0.17, 0.06]	[0.51, 1.10]	[−0.11, 0.10]	[−0.33, −0.01]	[−0.21, 0.11]	[−0.20, 0.14]	[−0.09, 0.22]
53	0.12	0.12	−0.07	−0.06	0.08	0.13	−0.08	0.16
	[−0.06, 0.30]	[−0.03, 0.26]	[−0.18, 0.05]	[−0.18, 0.07]	[−0.05, 0.22]	[−0.15, 0.40]	[−0.20, 0.05]	[−0.10, 0.42]

Table B1.2

Factor Loadings of the Exploratory Factor Analysis of Iteration 1 from Factor 9 to Factor 16

Items	F 9	F 10	F 11	F 12	F 13	F 14	F 15	F 16
1	−0.06	0.05	0.50*	−0.09	−0.08	0.19	0.05	−0.05
	[−0.25, 0.13]	[−0.09, 0.19]	[0.21, 0.80]	[−0.30, 0.13]	[−0.44, 0.29]	[−0.03, 0.40]	[−0.36, 0.45]	[−0.25, 0.16]
2	−0.05	0.03	−0.13	0.04	0.26	0.14	−0.08	−0.02
	[−0.26, 0.17]	[−0.11, 0.17]	[−0.35, 0.09]	[−0.15, 0.24]	[−0.05, 0.57]	[−0.05, 0.33]	[−0.31, 0.14]	[−0.22, 0.18]
3	0.08	> 0.000	−0.04	−0.11	0.11	−0.12	0.02	−0.002
	[−0.05, 0.22]	[−0.09, 0.09]	[−0.19, 0.10]	[−0.24, 0.03]	[−0.09, 0.30]	[−0.28, 0.03]	[−0.09, 0.14]	[−0.07, 0.07]
4	−0.17	−0.001	0.28	0.21	0.13	0.19	−0.16	0.17
	[−0.38, 0.04]	[−0.13, 0.13]	[−0.06, 0.63]	[−0.03, 0.44]	[−0.21, 0.47]	[−0.04, 0.42]	[−0.47, 0.16]	[−0.22, 0.55]
5	0.03	0.04	0.04	0.04	−0.05	−0.08	−0.01	0.09
	[−0.09, 0.16]	[−0.06, 0.13]	[−0.14, 0.21]	[−0.08, 0.16]	[−0.17, 0.08]	[−0.21, 0.06]	[−0.16, 0.14]	[−0.03, 0.21]
6	0.04	−0.05	0.02	−0.02	0.05	0.13*	−0.02	−0.08
	[−0.06, 0.13]	[−0.13, 0.04]	[−0.08, 0.11]	[−0.12, 0.08]	[−0.07, 0.18]	[0.02, 0.23]	[−0.12, 0.08]	[−0.18, 0.02]
7	0.03	0.04	−0.01	−0.05	−0.05	0.05	−0.01	−0.02
	[−0.02, 0.08]	[−0.02, 0.09]	[−0.08, 0.05]	[−0.14, 0.04]	[−0.14, 0.04]	[−0.02, 0.12]	[−0.07, 0.06]	[−0.06, 0.02]

8	−0.12	−0.003	−0.29	−0.01	0.31	−0.11	−0.02	0.23
	[−0.33, 0.08]	[−0.13, 0.13]	[−0.71, 0.12]	[−0.25, 0.22]	[−0.09, 0.71]	[−0.33, 0.11]	[−0.45, 0.41]	[−0.08, 0.54]
9	−0.22	−0.04	−0.19	0.13	0.31	0.06	0.03	−0.04
	[−0.50, 0.06]	[−0.20, 0.12]	[−0.47, 0.09]	[−0.14, 0.39]	[−0.01, 0.62]	[−0.17, 0.30]	[−0.32, 0.38]	[−0.32, 0.25]
10	0.09	0.03	−0.02	−0.13	0.08	−0.20	0.06	0.05
	[−0.06, 0.23]	[−0.10, 0.15]	[−0.19, 0.16]	[−0.33, 0.07]	[−0.27, 0.42]	[−0.43,0.03]	[−0.16, 0.28]	[−0.08, 0.18]
11	0.03	0.01	−0.02	0.05	−0.02	> 0.000	−0.004	−0.10
	[−0.03, 0.09]	[−0.04, 0.07]	[−0.07, 0.04]	[−0.03, 0.12]	[−0.08, 0.04]	[−0.05, 0.05]	[−0.06, 0.05]	[−0.32, 0.12]
12	0.01	−0.07	−0.04	0.04	−0.01	−0.01	−0.02	−0.04
	[−0.07, 0.10]	[−0.16, 0.02]	[−0.12, 0.05]	[−0.07, 0.15]	[−0.15, 0.13]	[−0.12, 0.10]	[−0.13, 0.08]	[−0.14, 0.06]
13	−0.03	0.01	−0.01	0.01	0.13	−0.14	0.14	−0.03
	[−0.20, 0.15]	[−0.12, 0.13]	[−0.29, 0.28]	[−0.17, 0.19]	[−0.14, 0.39]	[−0.30, 0.03]	[−0.05, 0.34]	[−0.22, 0.17]
14	0.75*	−0.02	−0.02	0.11	0.07	0.06	0.08	0.02
	[0.41, 1.08]	[−0.11, 0.07]	[−0.14, 0.11]	[−0.12, 0.35]	[−0.12, 0.25]	[−0.06, 0.19]	[−0.07, 0.22]	[−0.07, 0.10]
15	0.21*	0.12	−0.06	0.11	0.26*	0.21	−0.21	0.01
	[0.01, 0.41]	[−0.02, 0.26]	[−0.27, 0.14]	[−0.10, 0.33]	[0.05, 0.48]	[0.02, 0.39]	[−0.43, 0.02]	[−0.19, 0.21]

16	−0.01	−0.06	−0.35	−0.21	0.10	0.18	0.38	−0.08
	[−0.21, 0.19]	[−0.20, 0.08]	[−0.88, 0.18]	[−0.44, 0.02]	[−0.21, 0.42]	[−0.04, 0.40]	[−0.14, 0.90]	[−0.31, 0.14]
17	0.06	−0.01	−0.04	−0.02	0.18	0.08	0.12	0.02
	[−0.14, 0.25]	[−0.12, 0.10]	[−0.25, 0.17]	[−0.16, 0.12]	[−0.12, 0.48]	[−0.07, 0.23]	[−0.09, 0.33]	[−0.07, 0.10]
18	−0.49*	0.03	−0.02	0.17	0.14	0.02	0.16	0.02
	[−0.79, −0.19]	[−0.08, 0.15]	[−0.24, 0.19]	[−0.23, 0.38]	[−0.06, 0.39]	[−0.09, 0.18]	[−0.11, 0.39]	[−0.17, 0.20]
19	0.05	0.02	−0.09	−.06	0.16	0.08	−0.01	0.15
	[−0.19, 0.28]	[−0.13, 0.16]	[−0.31, 0.14]	[−0.27, 0.16]	[−0.08, 0.40]	[−0.10, 0.26]	[−0.24, 0.23]	[−0.07, 0.36]
20	0.01	0.08	0.01	0.02	0.004	0.02	−0.07	−0.06
	[−0.11, 0.13]	[−0.03, 0.19]	[−0.19, 0.20]	[−0.14, 0.18]	[−0.18, 0.19]	[−0.11, 0.15]	[−0.23, 0.10]	[−0.19, 0.07]
21	−0.01	1.18	−0.01	−0.001	0.003	0.003	0.02	−0.01
	[−0.04, 0.01]	[0.81, 1.55]	[−0.05, 0.03]	[−0.03, 0.03]	[−0.04, 0.04]	[−0.03, 0.04]	[−0.03, 0.06]	[−0.03, 0.02]
22	0.12	0.04	0.18	0.08	0.22	0.01	0.03	0.29*
	[−0.08, 0.32]	[−0.11, 0.18]	[−0.16, 0.53]	[−0.16, 0.32]	[−0.06, 0.50]	[−0.19, 0.22]	[−0.23, 0.29]	[0.02, 0.56]
23	0.04	0.15	0.23	−0.15	−0.09	0.01	0.15	0.06
	[−0.15, 0.23]	[−0.01, 0.30]	[−0.03, 0.49]	[−0.36, 0.06]	[−0.38, 0.21]	[−0.18, 0.21]	[−0.13, 0.43]	[−0.17, 0.28]

24	0.11	−0.09	0.25	−0.10	0.11	0.03	0.09	−0.08
	[−0.08, 0.30]	[−0.25, 0.07]	[−0.03, 0.53]	[−0.31, 0.10]	[−0.18, 0.39]	[−0.17, 0.23]	[−0.17, 0.34]	[−0.31, 0.14]
25	0.06	0.02	0.09	0.13	−0.05	0.25*	0.06	0.27
	[−0.14, 0.26]	[−0.12, 0.17]	[−0.15, 0.32]	[−0.09, 0.34]	[−0.26, 0.17]	[0.04, 0.46]	[−0.15, 0.26]	[−0.03, 0.57]
26	−0.03	0.16*	−0.01	0.23	−0.01	−0.01	0.25	0.19
	[−0.22, 0.17]	[0.02, 0.30]	[−0.31, 0.29]	[−0.002, 0.46]	[−0.23, 0.21]	[−0.20, 0.19]	[−0.01, 0.50]	[−0.05, 0.42]
27	0.05	0.11	0.12	0.10	−0.13	0.14	0.12	0.19
	[−0.17, 0.27]	[−0.03, 0.25]	[−0.09, 0.34]	[−0.13, 0.33]	[−0.38, 0.12]	[−0.06, 0.34]	[−0.13, 0.36]	[−0.09, 0.47]
28	0.13	−0.07	0.05	0.04	0.04	−0.04	0.40*	0.03
	[−0.09, 0.35]	[−0.22, 0.08]	[−0.44, 0.54]	[−0.19, 0.27]	[−0.33, 0.40]	[−0.23, 0.15]	[0.11, 0.68]	[−0.25, 0.31]
29	0.12	−0.03	−0.18	0.39*	0.04	−0.16	−0.09	0.03
	[−0.10, 0.34]	[−0.18, 0.12]	[−0.45, 0.09]	[0.15, 0.62]	[−0.15, 0.32]	[−0.38, 0.05]	[−0.30, 0.13]	[−0.20, 0.26]
30	0.18	0.07	−0.002	0.18*	0.60*	0.04	0.004	−0.04
	[−0.06, 0.41]	[−0.05, 0.18]	[−0.31, 0.31]	[0.01, 0.35]	[0.34, 0.85]	[−0.13, 0.21]	[−0.17, 0.18]	[−0.25, 0.18]
31	0.03	−0.08	−0.04	0.16	0.23	−0.09	0.37*	0.01
	[−0.19, 0.24]	[−0.23, 0.08]	[−0.58, 0.49]	[−0.07, 0.39]	[−0.001, 0.47]	[−0.31, 0.13]	[0.05, 0.69]	[−0.24, 0.25]

32	0.09	0.05	−0.03	0.81*	−0.04	0.07	0.12	−0.04
	[−0.09, 0.27]	[−0.06, 0.15]	[−0.13, 0.07]	[0.53, 1.10]	[−0.20, 0.12]	[−0.05, 0.19]	[−0.07, 0.30]	[−0.16, 0.08]
33	−0.11	−0.01	0.18	0.09	0.15	0.09	−0.02	−0.27
	[−0.35, 0.14]	[−0.16, 0.14]	[−0.08, 0.45]	[−0.16, 0.35]	[−0.17, 0.47]	[−0.12, 0.29]	[−0.24, 0.21]	[−0.59, 0.06]
34	0.09	0.10	0.25	−0.02	0.11	0.09	−0.25	0.02
	[−0.09, 0.28]	[−0.04, 0.25]	[−0.06, 0.56]	[−0.25, 0.21]	[−0.14, 0.35]	[−0.11, 0.29]	[−0.53, 0.04]	[−0.22, 0.26]
35	−0.09	−0.21*	−0.12	0.19	0.08	0.22*	0.20	0.02
	[−0.29, 0.12]	[−0.38, −0.03]	[−0.37, 0.14]	[−0.03, 0.42]	[−0.15, 0.32]	[0.02, 0.43]	[−0.06, 0.45]	[−0.19, 0.23]
36	−0.01	−0.02	0.13	0.21*	−0.01	0.12	−0.07	0.30
	[−0.21, 0.18]	[−0.16, 0.11]	[−0.16, 0.42]	[0.01, 0.40]	[−0.26, 0.25]	[−0.07, 0.32]	[−0.31, 0.17]	[−0.03, 0.62]
37	0.04	0.04	−0.01	−0.01	0.05	0.87*	0.01	0.002
	[−0.04, 0.13]	[−0.05, 0.12]	[−0.17, 0.16]	[−0.11, 0.10]	[−0.06, 0.16]	[0.61, 1.12]	[−0.06, 0.08]	[−0.05, 0.05]
38	−0.002	−0.09	0.02	0.11	−0.11	−0.03	0.15	0.02
	[−0.22, 0.21]	[−0.25, 0.07]	[−0.20, 0.24]	[−0.11, 0.32]	[−0.38, 0.16]	[−0.23, 0.17]	[−0.12, 0.42]	[−0.20, 0.24]
39	0.16	0.12	−0.09	0.05	0.14	0.07	0.17	0.05
	[−0.04, 0.36]	[−0.02, 0.27]	[−0.51, 0.32]	[−0.14, 0.24]	[−0.16, 0.44]	[−0.12, 0.25]	[−0.07, 0.41]	[−0.15, 0.24]

40	0.06	−0.21*	0.08	0.27	−0.24	−0.01	0.08	−0.10
	[−0.15, 0.28]	[−0.38, −0.04]	[−0.15, 0.32]	[−0.001, 0.54]	[−0.51, 0.03]	[−0.22, 0.19]	[−0.15, 0.30]	[−0.38, 0.19]
41	−0.24	0.02	0.05	0.43*	0.07	−0.13	0.15	−0.17
	[−0.48, 0.01]	[−0.10, 0.15]	[−0.18, 0.27]	[0.12, 0.73]	[−0.13, 0.26]	[−0.32, 0.05]	[−0.08, 0.37]	[−0.45, 0.10]
42	0.12	−0.09	0.52	−0.04	0.33	−0.20	0.15	−0.04
	[−0.07, 0.30]	[−0.23, 0.06]	[−0.11, 1.15]	[−0.25, 0.16]	[−0.14, 0.79]	[−0.44, 0.03]	[−0.23, 0.52]	[−0.28, 0.19]
43	0.11	0.04	0.19	0.40*	0.24	−0.19	−0.06	0.20
	[−0.10, 0.33]	[−0.09, 0.18]	[−0.10, 0.48]	[0.13, 0.67]	[−0.02, 0.49]	[−0.38, 0.01]	[−0.36, 0.23]	[−0.05, 0.45]
44	−0.07	−0.06	0.002	0.13	0.47*	0.14	0.07	0.14
	[−0.28, 0.14]	[−0.19, 0.07]	[−0.32, 0.33]	[−0.06, 0.32]	[0.11, 0.83]	[−0.06, 0.33]	[−0.16, 0.29]	[−0.07, 0.35]
45	0.01	−0.20*	0.06	0.09	−0.03	0.18	0.12	−0.05
	[−0.18, 0.20]	[−0.36, −0.04,]	[−0.15, 0.27]	[−0.13, 0.31]	[−0.28, 0.22]	[−0.02, 0.38]	[−0.09, 0.33]	[−0.26, 0.15]
46	−0.02	0.02	0.15	−0.04	−0.02	0.09	−0.07	0.12
	[0.17, −0.21]	[0.16, −0.11]	[0.38, −0.08]	[0.17, −0.24]	[0.19, −0.22]	[0.28, −0.10]	[0.15, −0.28]	[0.36, −0.13]
47	0.09	−0.01	0.08	0.28*	0.06	−0.09	−0.14	−0.09
	[−0.14, 0.32]	[−0.14, 0.12]	[−0.18, 0.34]	[0.05, 0.51]	[−0.15, 0.27]	[−0.26, 0.08]	[−0.37, 0.10]	[−0.34, 0.16]

48	0.09	0.14	0.12	0.15	0.12	−0.02	−0.05	−0.49*
	[−0.07, 0.26]	[0.01, 0.26]	[−0.10, 0.34]	[−0.04, 0.35]	[−0.08, 0.32]	[−0.18, 0.14]	[−0.21, 0.12]	[−0.81, −0.18]
49	0.06	0.12	0.10	0.16	−0.16	0.02	0.63*	−0.02
	[−0.14, 0.26]	[−0.02, 0.26]	[−0.23, 0.42]	[−0.11, 0.43]	[−0.41, 0.09]	[−0.15, 0.18]	[0.24, 1.02]	[−0.18, 0.15]
50	−0.07	0.07	0.12	0.09	0.06	0.06	0.43*	0.16
	[−0.27, 0.14]	[−0.07, 0.21]	[−0.23, 0.46]	[−0.13, 0.32]	[−0.19, 0.31]	[−0.13, 0.25]	[0.19, 0.66]	[−0.05, 0.37]
51	−0.13	−0.03	0.33*	−0.08	0.19	−0.03	0.14	0.02
	[−0.32, 0.07]	[−0.18, 0.12]	[0.05, 0.62]	[−0.31, 0.14]	[−0.15, 0.52]	[−0.18, 0.17]	[−0.21, 0.48]	[−0.22, 0.26]
52	−0.06	−0.03	0.01	−0.15	−0.02	−0.09	0.04	0.42*
	[−0.19, 0.07]	[−0.13, 0.06]	[−0.17, 0.18]	[−0.33, 0.03]	[−0.19, 0.14]	[−0.23, 0.05]	[−0.12, 0.20]	[0.16, 0.67]
53	−0.02	0.06	0.06	−0.08	0.82	0.08	−0.03	−0.04
	[−0.18, 0.14]	[−0.06, 0.17]	[−0.20, 0.32]	[−0.21, 0.06]	[0.44, 1.19]	[−0.06, 0.22]	[−0.17, 0.11]	[−0.18, 0.11]

Table B2

Factor Loadings of the Confirmatory Factor Analysis of Iteration 1

Items	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9	F 10	F 11	F 12	F 13	F 14	F 15	F 16
1	0.04															
2	0.75															
3	0.16															
4	0.21															
5		0.70														
6		0.89														
7		0.09														
8		0.48														
9			0.21													
10			0.06													
11			0.28													
12			0.22													
13				0.95												
14				0.64												

15	1.00				
16	0.24				
17		0.91			
18		0.40			
19		0.76			
20			0.92		
21			0.27		
22			0.43		
23			0.56		
24				0.49	
25				0.64	
26				0.47	
27				0.82	
28					0.65
29					−0.07
30					1.55

31	0.43					
32		1.07				
33		0.74				
34		0.51				
35			−0.26			
36			2.40			
37			1.05			
38				1.24		
39				2.43		
40					0.63	
41					0.76	
42					0.29	
43					0.47	
44						0.67
45						0.31
46						1.02

47	1.09		
48	0.63		
49	0.16		
50		0.00	
51		0.00	
52		8.29	
53			1.07

Notes. Confidence interval for factor loadings could not be calculated because standard errors were not available due to an unidentified model.

Table B3

Alpha and Beta Parameters of all Items in Iteration 1

Concepts	Items	Alpha	Beta
Population (Po)	1	−0.05	−3.22
	2	5.60	−0.60
	3	0.29	0.74
	4	0.46	−0.68
Randomness & Probability (RP)	5	1.88	0.06
	6	10.77	0.15
	7	0.10	−0.37
	8	0.70	−1.82
Decision, Risks and Consequences	9	1.54	−1.46
(DRC)	10	0.62	−1.76
	11	0.33	−4.00
	12	0.74	0.39
Research, Research methods &	13	0.61	0.51
Research designs (RRR)	14	0.67	0.04
	15	1.52	−0.69
	16	0.33	3.38
Distribution (Dis)	17	2.53	−0.76
	18	0.74	−0.39
	19	0.90	−0.23
Sample & Sampling (S)	20	1.01	−0.62

	21	0.64	1.20
	22	0.49	0.28
	23	0.78	−0.87
Outliers & Contamination (Co)	24	0.16	3.21
	25	0.69	−0.33
	26	0.62	1.38
	27	4.24	−0.53
Data, Observation & Summary	28	1.58	0.62
(DR)	29	−0.23	2.34
	30	0.74	−1.04
	31	0.45	2.16
Asymmetry (A)	32	0.41	3.14
	33	7.09	−0.63
	34	0.41	−0.89
Variability & Dispersion (VD)	35	0.35	2.33
	36	0.14	4.77
	37	2.67	0.34
Error & its Margins (EM)	38	0.69	1.14
	39	0.71	−0.33
Shared variance & Correlation (C)	40	0.84	1.49
	41	1.97	0.24
	42	0.28	4.43
	43	0.28	0.90

Central tendency (Ct)	44	0.79	−1.42
	45	0.67	1.26
	46	1.00	−1.04
Statistical Inference (SI)	47	1.08	0.43
	48	1.05	−0.03
	49	0.07	18.41
NHST & p-value (NP)	50	0.53	0.63
	51	1.21	0.30
	52	0.37	0.58
Modeling (M)	53	–	–

Table B4

Factor Loadings of the Exploratory Factor Analysis of Iteration 2

Items	F 1	F 2	F 3	F 4
1	0.12	0.63*	0.05	0.12
	[−0.11, 0.35]	[0.96, 0.30]	[0.50, −0.41]	[1.00, −0.77]
2	0.21	0.30	−0.20	−0.23
	[−0.10, 0.51]	[−0.08, 0.69]	[−0.72, 0.32]	[−0.83, 0.38]
3	1.46*	0.01	0.02	−0.01
	[0.10, 2.81]	[−0.04, 0.07]	[−0.10, 0.14]	[−0.07, 0.05]
4	0.02	0.15	−0.24	−0.06
	[−0.23, 0.26]	[−0.19, 0.49]	[−0.60, 0.13]	[−0.58, 0.47]
5	−0.02	0.04	−0.39	0.56*
	[−0.22, 0.19]	[−0.89, 0.97]	[−1.03, 0.25]	[0.03, 1.09]
6	0.05	−0.02	0.36	−0.34
	[−0.19, 0.28]	[−0.67, 0.62]	[−0.10, 0.82]	[−0.85, 0.16]
7	−0.07	−0.10	0.18	0.17
	[−0.31, 0.16]	[−0.45, 0.26]	[−0.20, 0.56]	[−0.27, 0.61]
8	−0.08	0.52	−0.34	0.45
	[−0.31, 0.15]	[−0.39, 1.42]	[−0.93, 0.25]	[−0.59, 1.48]
9	0.15	0.48*	0.28	0.20
	[−0.16, 0.45]	[0.07, 0.88]	[−0.15, 0.72]	[−0.38, 0.77]
10	0.21	−0.01	0.003	−0.01
	[−0.08, 0.51]	[−0.33, 0.31]	[−0.30, 0.31]	[−0.33, 0.32]

11	0.14	0.25	0.30	0.02
	[−0.15, 0.42]	[−0.17, 0.66]	[−0.09, 0.69]	[−0.41, 0.44]
12	0.05	0.71*	−0.04	−0.27
	[−0.16, 0.25]	[0.22, 1.20]	[−0.74, 0.66]	[−1.20, 0.65]
13	0.01	0.31	0.32	0.14
	[−0.22, 0.24]	[−0.01, 0.63]	[−0.02, 0.66]	[−0.21, 0.49]
14	0.14	−0.05	1.00*	0.11
	[−0.05, 0.32]	[−0.48, 0.38]	[0.68, 1.32]	[−0.71, 0.93]
15	0.13	0.28	−0.20	0.53
	[−0.11, 0.37]	[−0.55, 1.11]	[−0.75, 0.34]	[−0.17, 1.23]
16	−0.26	0.52	0.64	−0.13
	[−0.56, 0.04]	[−0.18, 1.23]	[−0.04, 1.32]	[−0.50, 0.24]
17	−0.08	−0.13	0.30	0.84*
	[−0.27, 0.12]	[−0.91, 0.64]	[−0.55, 1.16]	[0.13, 1.54]

Table B5

Factor Loadings of the Confirmatory Factor Analysis of Iteration 2

Items	F 1	F 2	F 3	F 4	F 5
1	61.54				
2	0.002				
3	0.002				
4	0.002				
5		0.83			
6		−0.40			
7		−0.01			
8		0.72			
9			0.44		
10			−0.03		
11			0.36		
12				0.38	
13				0.58	
14				0.65	
15					0.003
16					0.01
17					0.003

Notes. Confidence interval for factor loadings could not be calculated because standard errors were not available due to an unidentified model.

Table B6

Alpha and Beta Parameters of all Items in Iteration 2

Concepts	Items	Alpha	Beta
Population (Po)	1	0.48	−0.54
	2	1.25	−0.74
	3	0.81	−0.86
	4	0.29	−3.22
Randomness & Probability	5	11.33	−0.10
(RP)	6	−0.64	1.64
	7	0.17	−0.13
	8	1.50	−1.98
Sample & Sampling (S)	9	1.52	−1.69
	10	0.09	2.65
	11	0.46	−3.51
Variability & Dispersion	12	0.89	−1.49
(VD)	13	15.42	−0.07
	14	0.59	−2.20
Central tendency (Ct)	15	6.78	0.42
	16	−0.33	−3.22
	17	0.75	−1.97

Table B7

Factor Loadings of the Exploratory Factor Analysis of Iteration 3

Items	Factor 1	Factor 2	Factor 3	Factor 4
1	−0.08	0.31	−0.02	0.56*
	[−0.95, 0.79]	[−0.13, 0.76]	[−0.23, 0.19]	[0.21, 0.90]
2	0.20	0.39	0.22*	−0.13
	[−1.04, 1.43]	[−0.35, 1.14]	[0.02, 0.42]	[−0.38, 0.11]
3	0.12	0.40	−0.02	0.05
	[−1.00, 1.23]	[−0.29, 1.09]	[−0.24, 0.20]	[−0.26, 0.37]
4	−0.10	0.28	0.33*	0.19
	[−0.85, 0.65]	[−0.02, 0.58]	[0.11, 0.56]	[−0.11, 0.49]
5	−0.49*	0.05	−0.05	0.52*
	[−0.78, −0.19]	[−0.80, 0.89]	[−0.28, 0.17]	[0.20, 0.84]
6	0.03*	0.01	−0.08	−0.07
	[0.03, 0.31]	[−0.23, 0.26]	[−0.31, 0.15]	[−0.30, 0.15]
7	0.05	0.45	0.01	0.23
	[−1.15, 1.25]	[−0.29, 1.19]	[−0.30, 0.32]	[−0.26, 0.72]
8	0.54	0.23	0.12	0.04
	[−0.55, 1.62]	[−1.05, 1.51]	[−0.12, 0.35]	[−0.19, 0.27]
9	0.32	0.30	−0.02	0.22
	[−0.80, 1.44]	[−0.75, 1.36]	[−0.24, 0.20]	[−0.03, 0.47]
10	0.47*	−0.05	0.03	0.35*
	[0.08, 0.86]	[−1.04, 0.95]	[−0.20, 0.26]	[0.03, 0.66]

11	0.22	−0.21	0.47*	0.57*
	[−0.14, 0.57]	[−0.76, 0.35]	[0.25, 0.68]	[0.31, 0.85]
12	0.54	0.34	0.18	0.11
	[−0.88, 1.97]	[−1.20, 1.88]	[−0.05, 0.40]	[−0.12, 0.34]
13	0.86*	0.01	0.02	0.04
	[0.26, 1.46]	[−1.59, 1.60]	[−0.14, 0.20]	[−0.30, 0.37]
14	0.32	0.17	0.24*	0.14
	[−0.43, 1.06]	[−0.73, 1.06]	[0.04, 0.43]	[−0.06, 0.34]
15	0.08	0.08	0.92*	−0.05
	[−0.25, 0.40]	[−0.16, 0.32]	[0.79, 1.05]	[−0.13, 0.04]
16	−0.08	−0.04	1.05*	−0.02
	[−0.22, 0.07]	[−0.22, 0.15]	[0.94, 1.15]	[−0.09, 0.06]
17	0.12	0.56	0.09	−0.07
	[−1.44, 1.67]	[−0.21, 1.32]	[−0.11, 0.28]	[−0.42, 0.28]
18	0.11	0.34	0.14	0.30*
	[−0.94, 1.15]	[−0.34, 1.02]	[−0.06, 0.34]	[0.002, 0.60]
19	−0.23	0.33*	−0.10	0.29
	[−1.01, 0.56]	[0.03, 0.64]	[−0.31, 0.12]	[−0.09, 0.67]
20	−0.49	0.40	0.03	0.04
	[−1.27, 0.29]	[−0.29, 1.09]	[−0.17, 0.23]	[−0.40, 0.49]
21	0.17	0.02	0.02	0.03
	[−0.18, 0.53]	[−0.43, 0.48]	[−0.23, 0.26]	[−0.21, 0.27]
22	0.21	0.40	0.03	0.03

	[−1.02, 1.43]	[−0.44, 1.25]	[−0.18, 0.25]	[−0.24, 0.31]
23	0.11	0.04	−0.04	0.79*
	[−0.24, 0.47]	[−0.52, 0.60]	[−0.18, 0.11]	[0.63, 0.96]
24	−0.03	−0.02	−0.02	0.66*
	[−0.24, 0.18]	[−0.26, 0.22]	[−0.20, 0.16]	[0.49, 0.84]
25	0.31	0.47	−0.001	0.13
	[−1.26, 1.89]	[−0.76, 1.71]	[−0.20, 0.20]	[−0.17, 0.42]
26	0.18	0.38	−0.02	−0.14
	[−0.98, 1.34]	[−0.32, 1.07]	[−0.23, 0.18]	[−0.40, 0.12]
27	0.51	0.31	−0.07	0.21
	[−0.83, 1.83]	[−1.20, 1.82]	[−0.28, 0.14]	[−0.02, 0.45]
28	0.15	0.12	−0.09	0.35*
	[−0.38, 0.67]	[−0.50, 0.74]	[−0.29, 0.11]	[0.13, 0.58]
29	0.11	−0.29*	−0.07	0.47*
	[−0.54, 0.75]	[−0.54, −0.04]	[−0.27, 0.12]	[0.17, 0.76]
30	−0.003	0.21	0.04	0.60*
	[−0.65, 0.65]	[−0.28, 0.70]	[−0.17, 0.25]	[0.30, 0.89]
31	0.07	0.33	−0.14	0.29
	[−0.93, 1.07]	[−0.27, 0.92]	[−0.34, 0.07]	[−0.01, 0.60]
32	0.18	0.18	−0.07	0.39*
	[−0.49, 0.86]	[−0.55, 0.92]	[−0.25, 0.12]	[0.14, 0.64]
33	0.31	0.20	−0.11	0.20
	[−0.54, 1.16]	[−0.75, 1.15]	[−0.30, 0.09]	[−0.05, 0.43]

34	0.13	−0.14	−0.06	0.49*
	[−0.22, 0.48]	[−0.54, 0.25]	[−0.27, 0.15]	[0.24, 0.74]

Table B8

Factor Loadings of the Confirmatory Factor Analysis of Iteration 3

Items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1	0.65				
	[0.49, 0.80]				
2	0.43				
	[0.27, 0.60]				
3	0.43				
	[0.24, 0.61]				
4	0.46				
	[0.25, 0.67]				
5	0.08				
	[−0.17, 0.33]				
6	−0.05				
	[−0.24, 0.14]				
7	0.58				
	[0.33, 0.82]				
17	0.48				
	[0.31, 66]				
18	0.66				
	[0.51, 0.82]				
19	0.27				
	[0.09, 0.45]				

20	−0.02	
	[−0.22, 0.17]	
21	0.18	
	[−0.02, 0.37]	
22	0.50	
	[0.33, 0.67]	
8		0.66
		[0.46, 0.87]
9		0.60
		[0.41, 0.78]
10		0.63
		[0.42, 0.83]
11		0.70
		[0.55, 0.85]
12		0.83
		[0.65, 1.01]
13		0.72
		[0.54, 0.89]
14		0.59
		[0.45, 0.73]
15		0.98
		[0.91, 1.06]
16		0.90

	[0.82, 0.99]		
23		0.79	
		[0.67, 0.91]	
24		0.53	
		[0.38, 0.68]	
25		0.69	
		[0.52, 0.85]	
26		0.28	
		[0.10, 0.45]	
27		0.78	
		[0.58, 0.99]	
28		0.46	
		[0.30, 0.63]	
29			0.27
			[0.08, 0.46]
30			0.75
			[0.58, 0.92]
31			0.52
			[0.35, 0.68]
32			0.64
			[0.48, 0.81]
33			0.54
			[0.35, 0.74]

34	0.40
	[0.22, 0.59]

Table B9

Alpha and Beta Parameters of all Items of Iteration 3

Concepts	Items	Alpha	Beta
Population & Sample (PoS)	1	1.57	−0.62
	2	0.69	−0.72
	3	0.98	−1.28
	4	0.70	−2.32
	5	0.27	5.46
	6	0.07	7.98
	7	1.01	−2.74
	8	0.86	−1.21
	9	1.45	−0.95
	10	0.92	−0.41
	11	0.21	0.50
	12	0.20	5.56
Probability & Randomness (PR)	13	0.96	−2.51
	14	0.65	−2.85
	15	0.77	−2.94
	16	1.41	−1.57
	17	1.23	−1.05
	18	1.07	−2.41
	19	0.83	−1.08
	20	17.20	−0.62

	21	6.08	−0.58
Variability & Dispersion (VD)	22	4.39	−0.40
	23	1.58	−0.19
	24	0.84	−1.96
	25	0.37	0.72
	26	1.24	−2.10
	27	0.81	-0.86
Central tendency (Ct)	28	0.63	0.53
	29	1.83	−1.29
	30	0.85	−0.32
	31	1.65	−0.82
	32	0.98	−0.77
	33	0.63	0.83

Table B10

Exploratory Factor Analysis of Iteration 4

Items	F1	F2	F3	F4	F5
1	0.41*	0.39	−0.09	0.19	−0.03
	[0.02, 0.80]	[−0.02, 0.80]	[−0.31, 0.14]	[−0.11, 0.48]	[−0.32, 0.26]
2	0.41*	0.27	−0.06	−0.06	−0.20
	[0.08, 0.73]	[−0.06, 0.60]	[−0.34, 0.23]	[−0.35, 0.23]	[−0.50, 0.10]
3	0.51*	0.26	−0.25*	−0.03	−0.02
	[0.13, 0.89]	[−0.17, 0.70]	[−0.48, −0.03]	[−0.30, 0.24]	[−0.38, 0.34]
4	0.25	0.08	−0.12	−0.30	−0.33
	[−0.04, 0.54]	[−0.22, 0.39]	[−0.47, 0.24]	[−0.63, 0.03]	[−0.67, 0.003]
5	−0.17	−0.01	0.15	−0.06	−0.001
	[−0.46, 0.13]	[−0.31, 0.29]	[−0.11, 0.41]	[−0.35, 0.23]	[−0.30, 0.30]
6	0.35	0.06	0.41	−0.05	0.14
	[−0.09, 0.79]	[−0.26, 0.39]	[−0.03, 0.85]	[−0.47, 0.38]	[−0.21, 0.49]
7	0.49*	0.25	0.11	−0.12	−0.002
	[0.10, 0.87]	[−0.12, 0.62]	[−0.14, 0.37]	[−0.41, 0.16]	[−0.34, 0.34]
8	0.38	0.34	−0.17	0.16	0.45*
	[−0.26, 1.02]	[−0.14, 0.82]	[−0.40, 0.07]	[−0.07, 0.39]	[0.14, 0.77]
9	0.15	0.25	0.08	0.03	0.09
	[−0.23, 0.52]	[−0.04, 0.54]	[−0.20, 0.36]	[−0.23, 0.29]	[−0.17, 0.34]
10	−0.23	0.43	0.16	−0.01	0.07
	[−0.67, 0.22]	[0.19, 0.68]	[−0.14, 0.45]	[−0.30, 0.27]	[−0.24, 0.39]

11	0.23	−0.05	0.11	−0.21	0.71*
	[−0.44, 0.91]	[−0.38, 0.28]	[−0.13, 0.34]	[−0.47, 0.05]	[0.31, 1.12]
12	0.53*	0.29	−0.01	0.13	−0.04
	[0.19, 0.88]	[−0.13, 0.71]	[−0.23, 0.21]	[−0.15, 0.40]	[−0.36, 0.29]
13	0.93*	−0.09	0.15	0.23	−0.36
	[0.40, 1.46]	[−0.43, 0.25]	[−0.04, 0.35]	[−0.12, 0.58]	[−0.71, −0.001]
14	0.18	0.58*	−0.10	0.08	0.01
	[−0.35, 0.72]	[0.18, 0.97]	[−0.34, 0.15]	[−0.23, 0.39]	[−0.28, 0.29]
15	−0.39	0.62*	0.29*	−0.23	−0.27
	[−0.84, 0.06]	[0.18, 1.06]	[0.02, 0.56]	[−0.45, 0.001]	[−0.75, 0.21]
16	0.56*	−0.01	0.06	0.37*	0.23
	[0.25, 0.88]	[−0.38, 0.37]	[−0.19, 0.32]	[0.02, 0.72]	[−0.19, 0.66]
17	0.89*	−0.02	0.20*	0.02	0.21
	[0.60, 1.18]	[−0.42, 0.38]	[0.05, 0.35]	[−0.26, 0.31]	[−0.33, 0.74]
18	0.69*	0.13	0.08	−0.12	0.23
	[0.24, 1.14]	[−0.30, 0.55]	[−0.13, 0.28]	[−0.39, 0.15]	[−0.25, 0.70]
19	−0.04	0.20	−0.20	−0.37	0.04
	[−0.51, 0.42]	[−0.07, 0.46]	[−0.53, 0.13]	[−0.66, −0.08]	[−0.28, 0.35]
20	0.07	0.004	1.09	0.02	0.07
	[−0.10, 0.23]	[−0.15, 0.15]	[0.88, 1.30]	[−0.11, 0.15]	[−0.04, 0.18]
21	−0.03	0.06	0.73*	0.22*	−0.08
	[−0.32, 0.25]	[−0.21, 0.32]	[0.52, 0.93]	[0.01, 0.43]	[−0.31, 0.15]
22	0.16	0.14	0.12	0.82*	0.06

	[−0.13, 0.45]	[−0.15, 0.44]	[−0.05, 0.28]	[0.61, 1.04]	[−0.11, 0.22]
23	0.07	0.04	0.12	0.84*	0.04
	[−0.31, 0.45]	[−0.16, 0.23]	[−0.06, 0.30]	[0.63, 1.05]	[−0.12, 0.19]
24	0.31*	0.06	0.004	−0.06	0.06
	[0.04, 0.58]	[−0.24, 0.35]	[−0.27, 0.28]	[−0.33, 0.21]	[−0.24, 0.35]
25	0.51	0.24	−0.10	−0.20	0.22
	[−0.06, 1.08]	[−0.18, 0.66]	[−0.37, 0.18]	[−0.47, 0.06]	[−0.17, 0.61]
26	0.13	0.76*	−0.01	0.19	0.09
	[−0.49, 0.76]	[0.18, 1.35]	[−0.22, 0.21]	[−0.23, 0.61]	[−0.23, 0.40]
27	0.05	0.11	−0.05	0.21	0.03
	[−0.25, 0.34]	[−0.15, 0.36]	[−0.32, 0.23]	[−0.04, 0.46]	[−0.22, 0.28]
28	−0.20	0.53*	0.01	0.15	0.01
	[−0.70, 0.31]	[0.30, 0.77]	[−0.27, 0.28]	[−0.11, 0.41]	[−0.34, 0.36]
29	−0.49	0.02	−0.36	0.58*	−0.11
	[−0.97, −0.01]	[−0.20, 0.25]	[−0.74, 0.02]	[0.20, 0.95]	[−0.52, 0.31]
30	0.05	0.31	−0.24	0.26*	0.22
	[−0.37, 0.48]	[−0.02, 0.65]	[−0.48, −0.002]	[0.53, 0.0004]	[−0.03, 0.47]
31	−0.24	0.01	0.11	0.16	0.85*
	[−0.80, 0.33]	[−0.18, 0.20]	[−0.14, 0.36]	[−0.09, 0.41]	[0.52, 1.18]
32	0.02	0.31*	0.21	0.10	0.11
	[−0.34, 0.39]	[0.05, 0.56]	[−0.08, 0.49]	[−0.14, 0.34]	[−0.15, 0.36]
33	0.36*	0.24	0.14	0.27*	0.01
	[0.10, 0.63]	[−0.08, 0.56]	[−0.09, 0.37]	[0.03, 0.50]	[−0.27, 0.28]

Table B11

Confirmatory Factor Analysis of Iteration 4

Items	F1	F2	F3	F4
1	0.63*			
	[0.43, 0.82]			
2	0.36*			
	[0.14, 0.59]			
3	0.46*			
	[0.25, 0.67]			
4	−0.03			
	[−0.33, 0.27]			
5	−0.11			
	[−0.36, 0.14]			
6	0.61*			
	[0.31, 0.91]			
7	0.56*			
	[0.33, 0.78]			
8	0.69*			
	[0.52, 0.86]			
9	0.35*			
	[0.14, 0.56]			
10	0.15			
	[−0.09, 0.38]			

11	0.44*		
	[0.29, 0.59]		
12	0.68		
	[0.50, 0.85]		
13		0.82*	
		[0.59, 1.04]	
14		0.49*	
		[0.22, 0.76]	
15		−0.08	
		[−0.30, 0.15]	
16		0.80*	
		[0.60, 1.01]	
17		1.04*	
		[0.89, 1.18]	
18		0.76*	
		[0.60, 0.92]	
19		−0.23	
		[−0.51, 0.05]	
20		0.73*	
		[0.53, 0.93]	
21		0.57*	
		[0.38, 0.77]	
22			0.92*

	[0.80, 1.04]	
23	0.78*	
	[0.63, 0.92]	
24	0.29*	
	[0.07, 0.50]	
25	0.54*	
	[0.34, 0.74]	
26	0.69*	
	[0.38, 1.00]	
27	0.24*	
	[0.003, 0.47]	
28		0.23*
		[0.01, 0.45]
29		−0.35*
		[−0.62, −0.07]
30		0.34
		[0.11, 0.57]
31		0.17
		[−0.06, 0.40]
32		0.38*
		[0.17, 0.59]
33		0.67
		[0.43, 0.91]

Table B12

Alpha and Beta Parameters of all Items of Iteration 4

Concepts	Items	Alpha	Beta
Population & Sampling (PoS)	1	1.60	−0.38
	2	0.70	−0.66
	3	1.12	−1.05
	4	0.25	5.22
	5	−0.26	−4.13
	6	0.96	−3.49
	7	1.16	−1.63
	8	1.07	−1.16
	9	0.66	−0.44
	10	0.24	2.62
	11	0.50	2.34
	12	1.95	−0.71
Randomness & Probability (RP)	13	2.71	−1.91
	14	0.41	−4.66
	15	−0.31	2.08
	16	2.19	−1.80
	17	15.29	−1.48
	18	2.18	−1.36
	19	0.58	−1.16
	20	1.47	−1.89

	21	0.88	−1.58
Variability & Dispersion (VD)	22	16.61	−0.61
	23	2.63	−0.62
	24	0.18	2.08
	25	0.27	−0.33
	26	0.84	−3.87
	27	0.52	1.39
Central tendency (Ct)	28	1.09	−0.69
	29	0.48	−0.58
	30	0.35	−2.52
	31	0.81	−2.05
	32	1.07	0.24
	33	0.87	−0.88

Table B13

The 3-Factor Model of the Exploratory Factor Analysis of Iteration 5

Items	F1	F2	F3
1	0.57*	0.14	−0.09
	[0.43, 0.71]	[−0.04, 0.31]	[−0.24, 0.06]
2	0.17	0.32*	−0.06
	[-0.01, 0.35]	[0.12, 0.52]	[−0.25, 0.13]
3	0.30*	0.49*	>0.000
	[0.12, 0.48]	[0.30, 0.67]	[−0.16, 0.16]
4	0.15	0.52	−0.13
	[−0.05, 0.35]	[0.33, 0.72]	[−0.32, 0.06]
5	0.07	0.21*	0.01
	[−0.11, 0.25]	[0.03, 0.40]	[−0.17, 0.19]
6	0.37*	0.11	−0.20*
	[0.21, 0.53]	[−0.10, 0.33]	[−0.02, −0.37]
7	0.39*	0.22*	−0.15
	[0.23, 0.55]	[0.02, 0.42]	[−0.32, 0.02]
8	0.25*	0.41*	0.06
	[0.08, 0.42]	[0.23, 0.59]	[−0.10, 0.23]
9	0.21*	−0.30*	−0.07
	[0.05, 0.37]	[−0.48, −0.12]	[−0.26, 0.12]
10	0.20*	0.41*	0.14
	[0.02, 0.39]	[0.22, 0.60]	[−0.05, 0.34]

11	0.09	0.39*	0.48*
	[−0.06, 0.23]	[0.22, 0.57]	[0.31, 0.65]
12	0.01	−0.04	0.92*
	[−0.08, 0.10]	[−0.14, 0.07]	[0.76, 1.07]
13	0.01	0.02	0.87*
	[−0.09, 0.12]	[−0.12, 0.16]	[0.71, 1.03]
14	0.94*	−0.06	0.06
	[0.85, 1.03]	[−0.14, 0.01]	[−0.02, 0.14]
15	0.86*	0.02	0.01
	[0.76, 0.96]	[−0.09, 0.12]	[−0.09, 0.11]
16	−0.15*	0.64*	−0.05
	[−0.28, −0.01]	[0.49, 0.79]	[−0.19, 0.10]
17	−0.08	0.53*	0.06
	[−0.23, 0.07]	[0.36, 0.69]	[−0.12, 0.24]
18	0.30*	0.02	−0.08
	[0.14, 0.46]	[−0.17, 0.21]	[−0.25, 0.10]
19	0.20*	0.01	0.01
	[0.04, 0.36]	[−0.18, 0.20]	[−0.17, 0.19]
20	0.34*	0.07	0.13
	[0.19, 0.50]	[−0.13, 0.27]	[−0.04, 0.31]
21	0.31*	0.04	0.01
	[0.15, 0.46]	[−0.16, 0.23]	[−0.17, 0.19]
22	−0.37*	0.24*	0.17

	[−0.53, −0.22]	[0.06, 0.45]	[−0.01, 0.35]
23	0.21*	0.25*	−0.18*
	[0.04, 0.38]	[0.07, 0.44]	[−0.35, −0.001]
24	0.14	0.25	−0.04
	[−0.02, 0.31]	[0.07, 0.42]	[−0.22, 0.13]
25	−0.04	0.40*	0.27
	[−0.17, 0.10]	[0.23, 0.58]	[0.08, 0.45]
26	−0.08	0.58*	0.13
	[−0.23, 0.07]	[0.38, 0.77]	[−0.05, 0.30]

Table B14

The 4-Factor Model of the Exploratory Factor Analysis of Iteration 5

Items	F1	F2	F3	F4
1	0.31*	0.47*	−0.02	−0.05
	[0.11, 0.50]	[0.30, 0.64]	[−0.17, 0.13]	[−0.24, 0.15]
2	0.49*	−0.04	0.05	0.04
	[0.28, 0.70]	[−0.22, 0.15]	[−0.14, 0.24]	[−0.17, 0.26]
3	0.34*	0.24*	0.04	0.32*
	[0.10, 0.57]	[0.06, 0.43]	[−0.14, 0.21]	[0.10, 0.53]
4	0.65*	−0.11	−0.01	0.18
	[0.41, 0.90]	[−0.26, 0.04]	[−0.17, 0.15]	[−0.07, 0.42]
5	0.19	0.01	0.-4	0.12
	[−0.03, 0.40]	[−0.17, 0.20]	[−0.14, 0.22]	[−0.07, 0.31]
6	0.33*	0.24*	−0.11	−0.08
	[0.12, 0.53]	[0.05, 0.42]	[−0.29, 0.07]	[−0.32, 0.15]
7	0.70*	0.09	0.01	−0.18*
	[0.51, 0.90]	[−0.08, 0.25]	[−0.11, 0.14]	[−0.36, 0.01]
8	0.36*	0.15	0.12	0.22*
	[0.13, 0.58]	[−0.03, 0.32]	[−0.05, 0.28]	[0.02, 0.43]
9	0.25*	0.01	0.05*	−0.50*
	[0.02, 0.48]	[−0.16, 0.17]	[−0.12, 0.22]	[−0.68, −0.31]
10	0.27*	0.16	0.18	0.27*
	[0.03, 0.51]	[−0.03, 0.34]	[−0.02, 0.37]	[0.06, 0.48]

11	0.17	0.07	0.49*	0.31*
	[−0.04, 0.38]	[−0.09, 0.24]	[0.32, 0.66]	[0.12, 0.50]
12	0.004	−0.03	0.98*	−0.07*
	[−0.09, 0.10]	[−0.10, 0.04]	[0.83, 1.12]	[−0.13, −0.01]
13	−0.03	0.02	0.83*	0.05
	[−0.18, 0.11]	[−0.09, 0.13]	[0.67, 1.00]	[−0.10, 0.20]
14	−0.05	1.00*	0.02	−0.02
	[−0.14, 0.04]	[0.90, 1.10]	[−0.05, 0.08]	[−0.09, 0.05]
15	0.03	0.84*	−0.04	0.05
	[−0.13, 0.18]	[0.72, 0.97]	[−0.13, 0.06]	[−0.05, 0.16]
16	0.23	−0.14	−0.06	0.56*
	[−0.01, 0.46]	[−0.27, 0.001]	[−0.21, 0.08]	[0.36, 0.76]
17	0.04	0.02	0.02	0.53*
	[−0.16, 0.24]	[−0.14, 0.18]	[−0.15, 0.19]	[0.36, 0.70]
18	0.17	0.22*	−0.04	−0.08
	[−0.03, 0.37]	[0.04, 0.40]	[−0.22, 0.14]	[−0.28, 0.13]
19	0.10	0.16	0.04	−0.05
	[−0.10, 0.30]	[−0.02, 0.33]	[−0.14, 0.22]	[−0.24, 0.14]
20	0.08	0.33*	0.15	0.02
	[−0.13, 0.29]	[0.16, 0.50]	[−0.03, 0.32]	[−0.18, 0.22]
21	0.10	0.27*	0.02	−0.02
	[−0.10, 0.30]	[0.10, 0.44]	[−0.16, 0.21]	[−0.21, 0.18]
22	−0.09	−0.28*	0.13	0.32*

	[−0.29, 0.11]	[−0.45, −0.11]	[−0.05, 0.31]	[0.12, 0.51]
23	0.18	0.18*	−0.16	0.16
	[−0.01, 0.38]	[0.01, 0.36]	[−0.33, 0.02]	[−0.04, 0.36]
24	−0.01	0.22*	−0.08	0.28*
	[−0.23, 0.20]	[0.05, 0.40]	[−0.25, 0.09]	[0.09, 0.46]
25	−0.09	0.10	0.20*	0.49*
	[−0.30, 0.12]	[−0.05, 0.26]	[0.03, 0.38]	[0.32, 0.65]
26	−0.03	0.07	0.05	0.65*
	[−0.21, 0.15]	[−0.07, 0.21]	[−0.10, 0.20]	[0.47, 0.83]

Table B15

The Confirmatory Factor Analysis of Iteration 5

Items	F1	F2	F3	F4	F5
1	0.61*				
	[0.47, 0.75]				
2	0.40*				
	[0.24, 0.57]				
3	0.67*				
	[0.51, 0.83]				
4	0.50*				
	[0.30, 0.70]				
5	0.25*				
	[0.08, 0.42]				
6	0.34*				
	[0.17, 0.51]				
7	0.49*				
	[0.34, 0.64]				
8	0.60*				
	[0.43, 0.76]				
9	−0.06				
	[−0.24, 0.12]				
10		0.56*			
		[0.41, 0.74]			

11	0.77*		
	[0.64, 0.90]		
12	0.84*		
	[0.73, 0.95]		
13	0.83*		
	[0.71, 0.95]		
14		0.90*	
		[0.79, 1.01]	
15		0.88*	
		[0.78, 0.99]	
16		0.33*	
		[0.17, 0.48]	
17		0.33*	
		[0.17, 0.48]	
18		0.25*	
		[0.10, 0.40]	
19			0.23*
			[0.04, 0.41]
20			0.51*
			[0.28, 0.74]
21			0.37*
			[0.17, 0.57]
22			−0.10

	[−0.30, 0.09]	
23		0.31*
		[0.13, 0.49]
24		0.33*
		[0.16, 0.51]
25		0.56*
		[0.39, 0.72]
26		0.56*
		[0.38, 0.73]

Table B16

Alpha and Beta Parameters of all Items of Iteration 5

Concepts	Items	Alpha	Beta
Population & Sampling (PoS)	1	1.08	−0.92
	2	0.90	−1.62
	3	1.19	−1.41
	4	1.11	2.29
	5	0.42	−1.83
	6	0.71	−1.32
	7	1.67	−0.56
	8	0.95	−1.63
	9	0.22	4.36
Randomness & Probability (RP)	10	0.62	−2.53
	11	1.48	−1.42
	12	4.00	−0.53
	13	3.07	−0.30
Variability & Dispersion (VD)	14	2.86	−0.43
	15	10.53	−0.62
	16	0.09	2.54
	17	0.12	−2.97
	18	0.49	0.67
Central tendency (Ct)	19	0.68	−0.47
	20	0.80	−0.90

	21	0.47	−0.81
	22	−0.35	−2.03
Shared Variance & Correlation (Co)	23	0.50	−0.02
	24	0.45	−0.16
	25	0.94	0.35
	26	1.57	0.48

Figures of the Appendix B

Figure B1

Item Characteristics Curves of items 1 to 4 in Factor 1 (Population) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability.

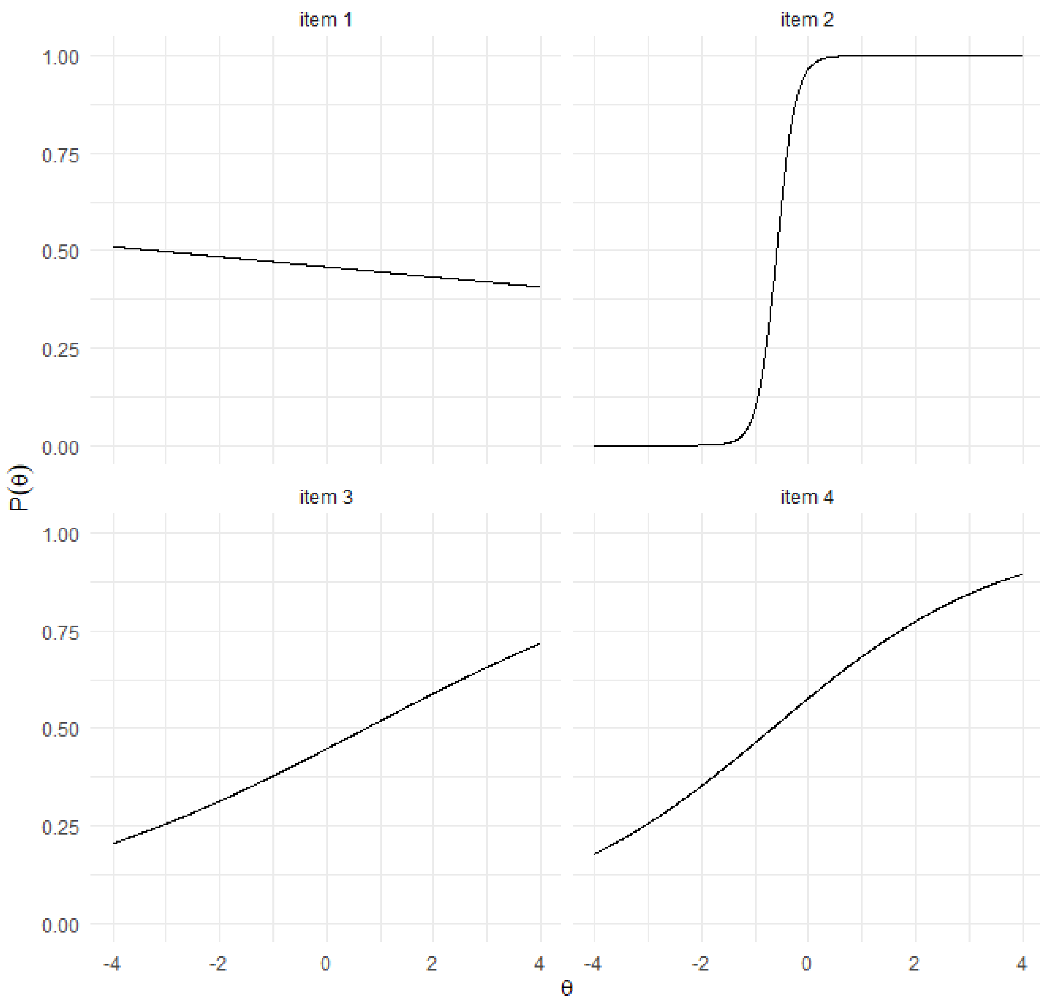


Figure B2

Item Characteristics Curves of items 5 to 8 in Factor 2 (Randomness & Probability) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 5, item 2 = item 6, item 3 = item 7 and item 4 = item 8.

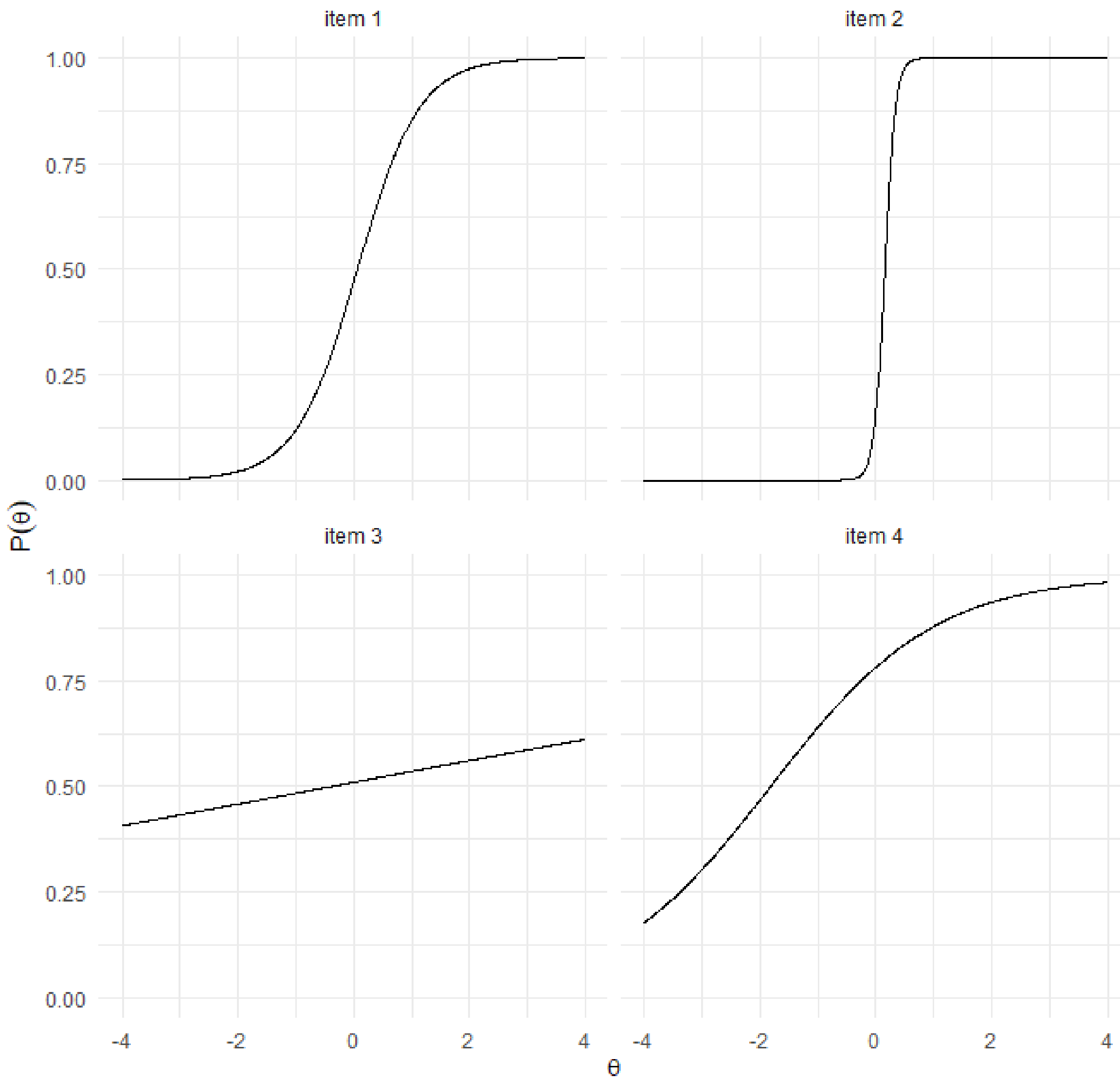


Figure B3

Item Characteristics Curves of items 9 to 12 in Factor 3 (Decision, Risks and Consequences) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 9, item 2 = item 10, item 3 = item 11 and item 4 = item 12.

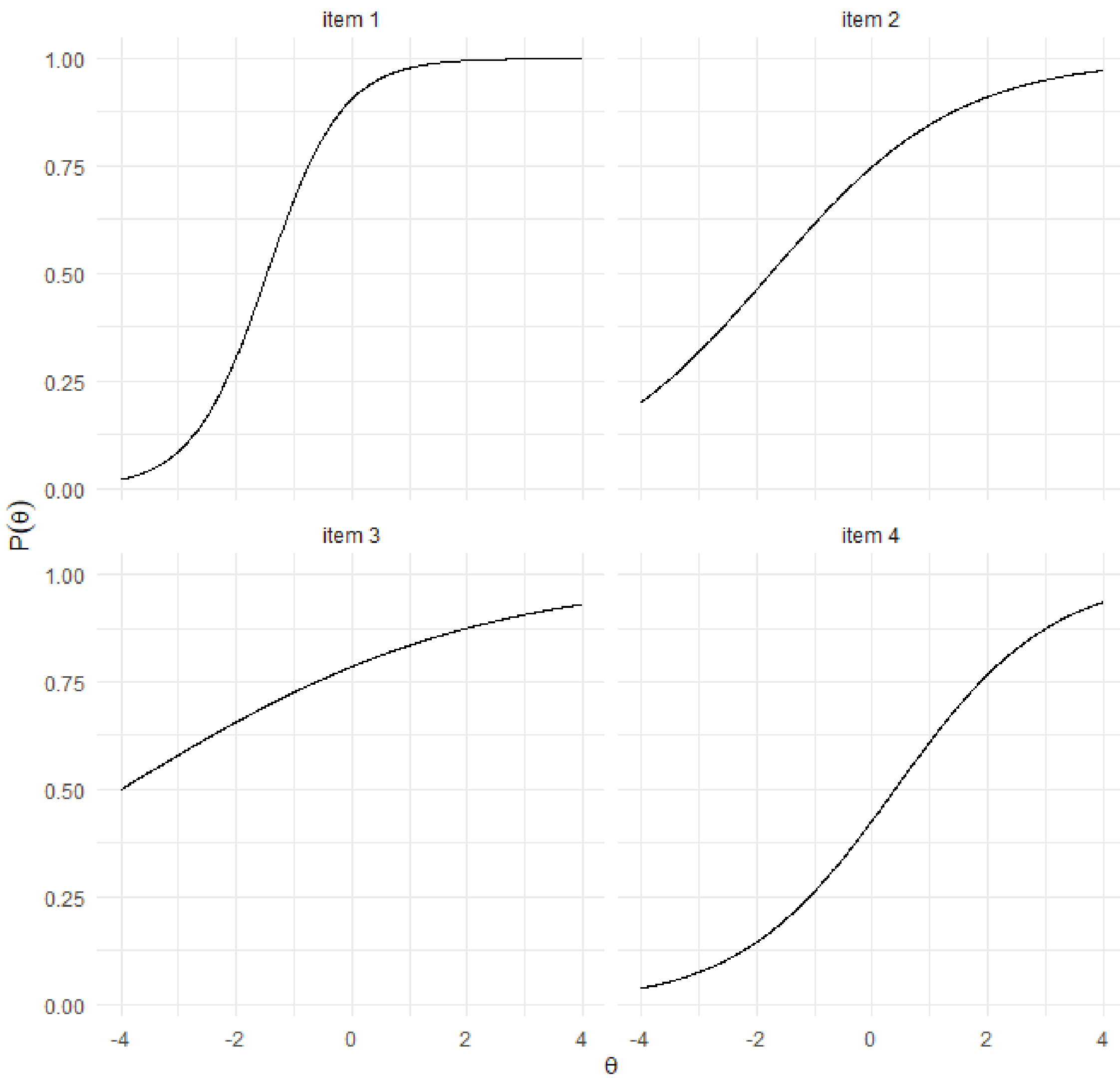


Figure B4

Item Characteristics Curves of items 13 to 16 in Factor 4 (Research, Research methods & Research designs) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 13, item 2 = item 14, item 3 = item 15 and item 4 = item 16.

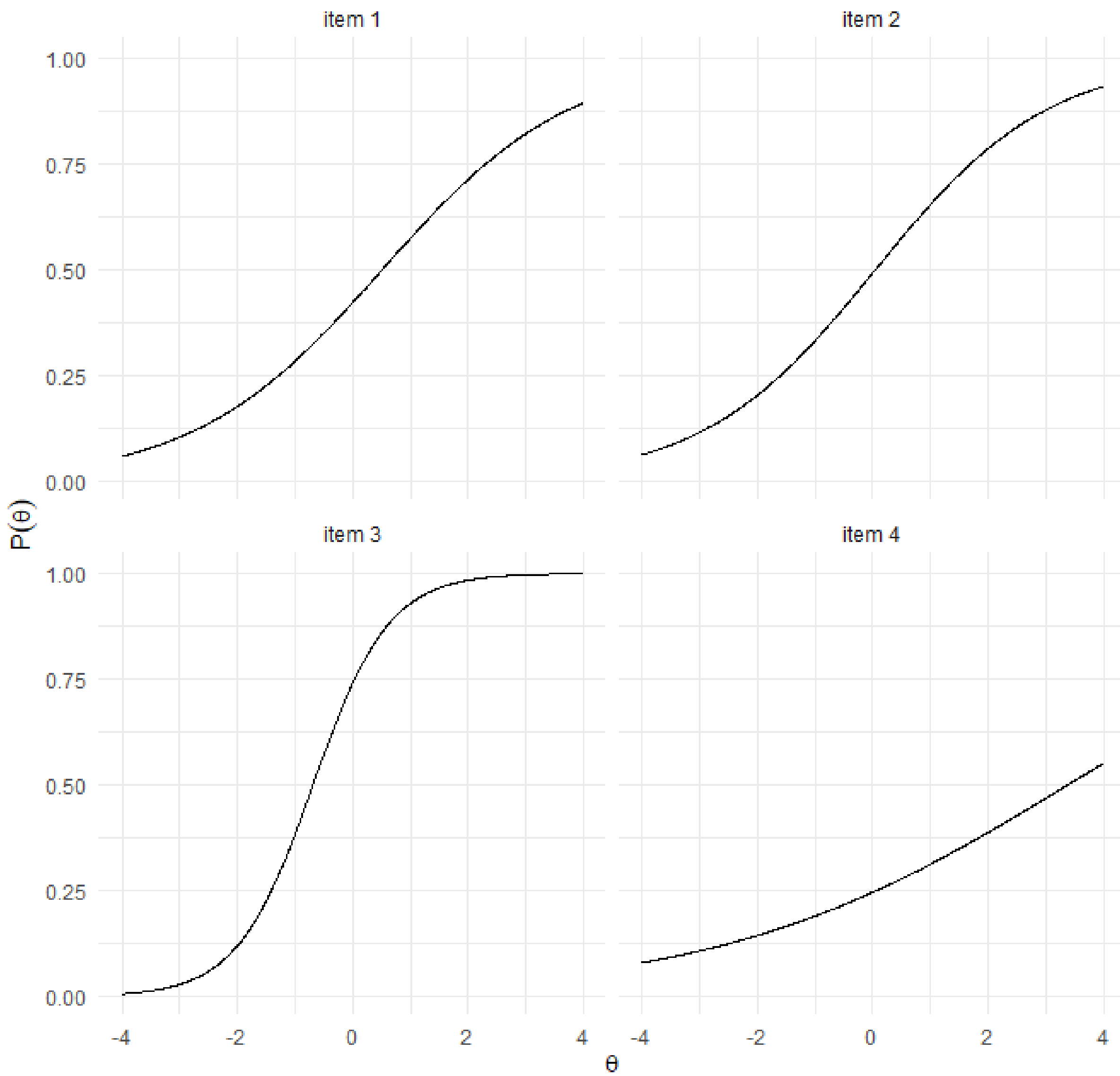


Figure B5

Item Characteristics Curves of items 17 to 19 in Factor 5 (Distribution) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 17, item 2 = item 18 and item 3 = item 19.

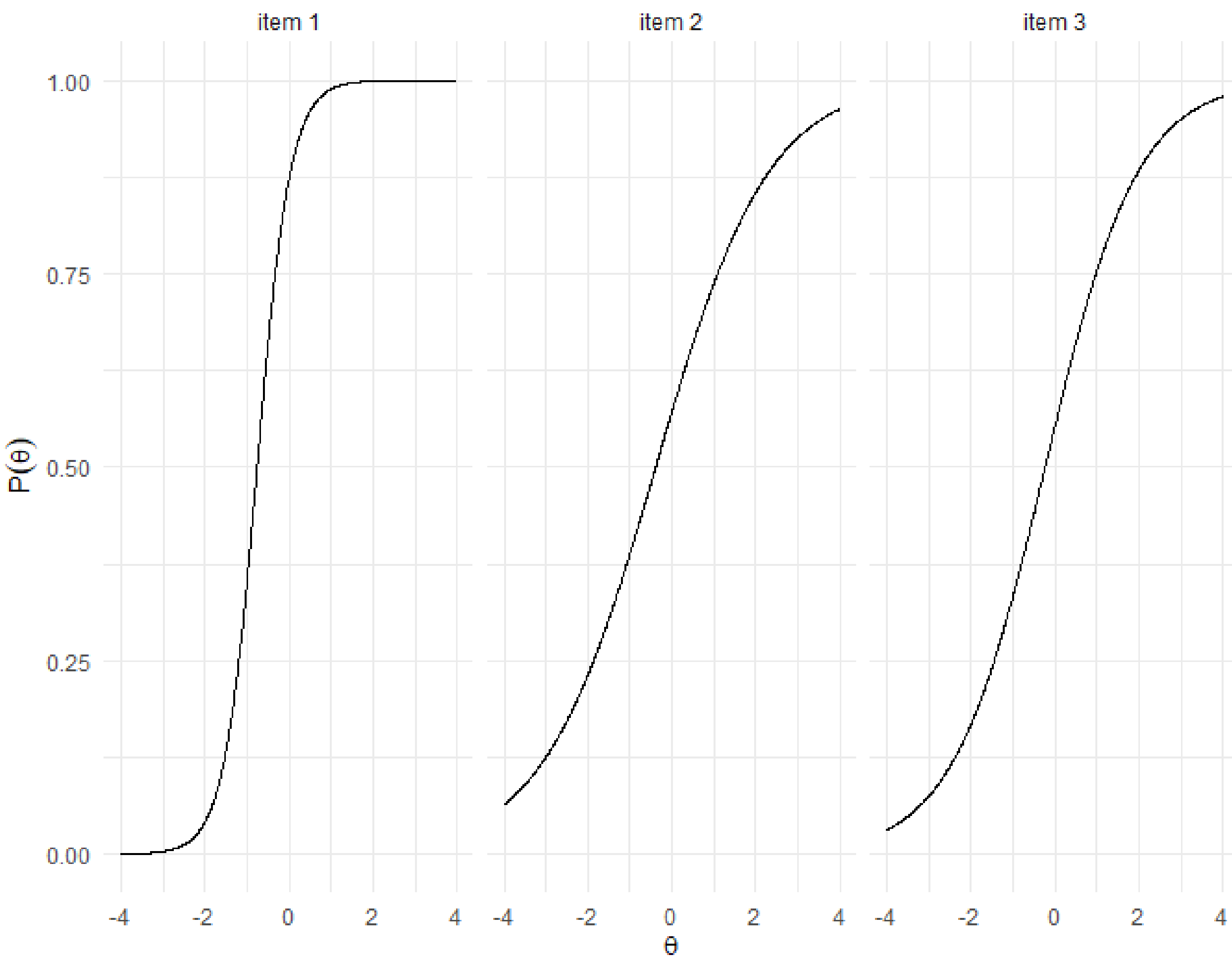


Figure B6

Item Characteristics Curves of items 20 to 23 in Factor 6 (Sample & Sampling) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 20, item 2 = item 21, item 3 = item 22 and item 4 = item 23.

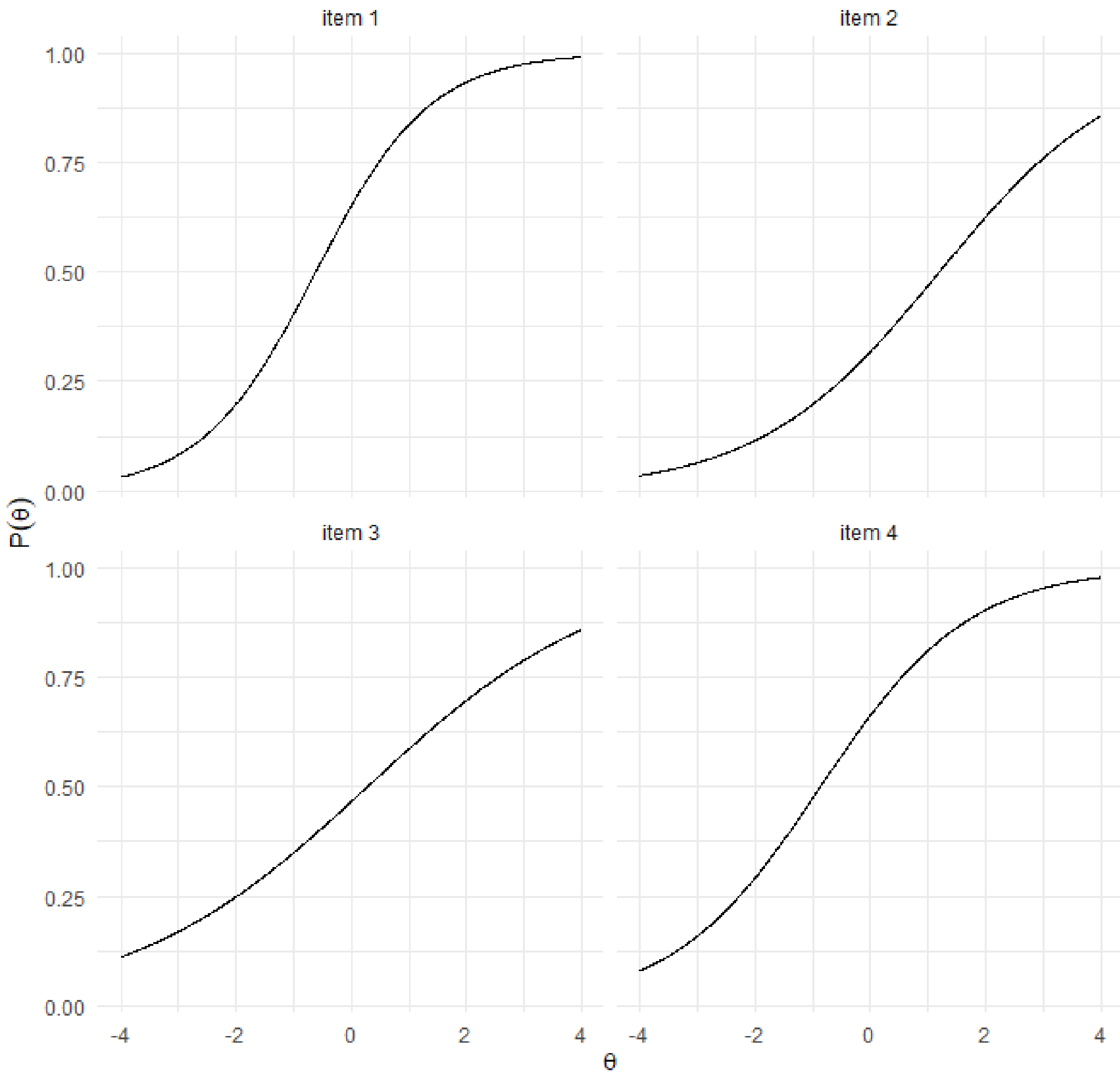


Figure B7

Item Characteristics Curves of items 24 to 27 in Factor 7 (Outliers & Contamination) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 24, item 2 = item 25, item 3 = item 26 and item 4 = item 27.

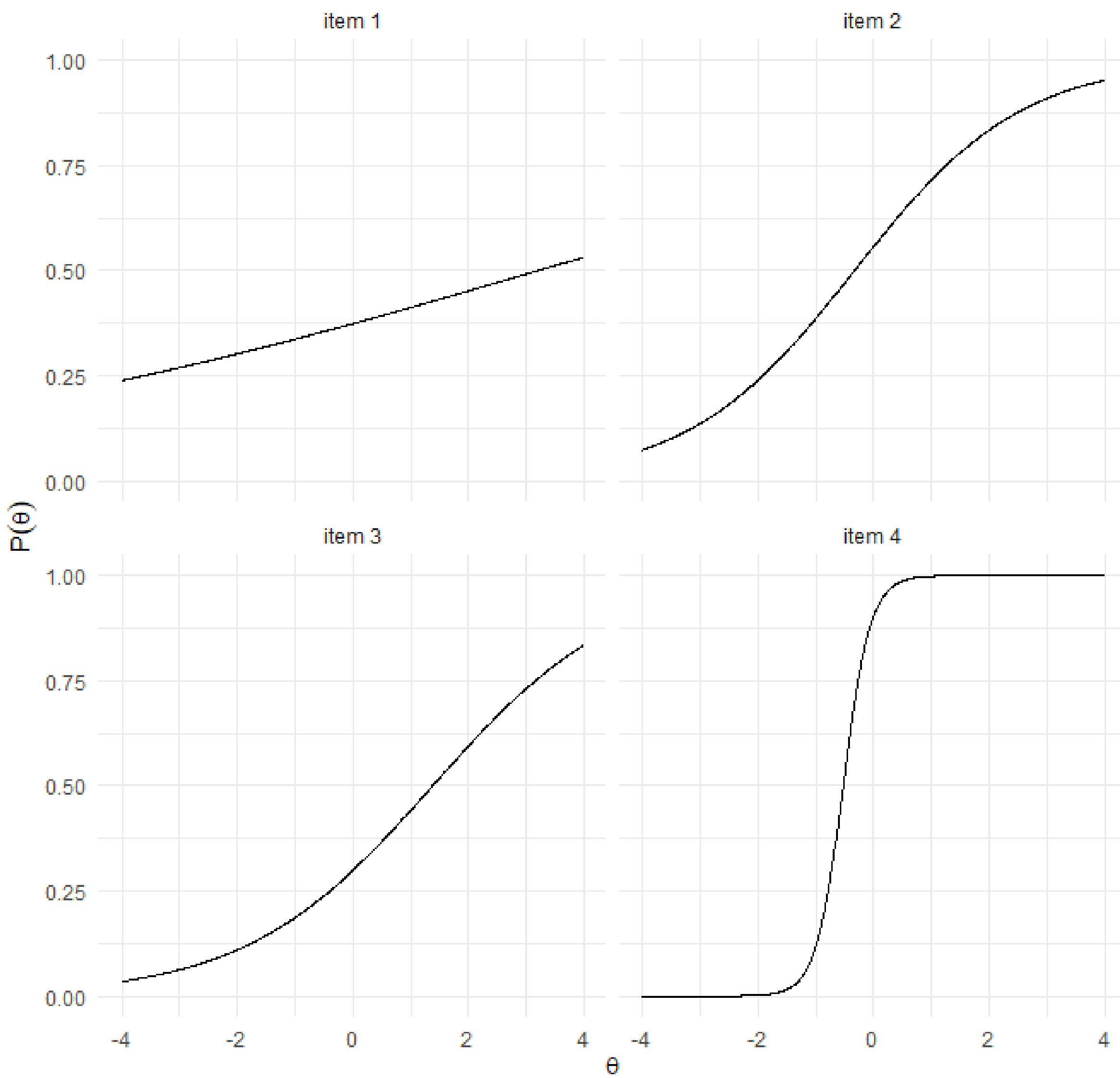


Figure B8

Item Characteristics Curves of items 28 to 31 in Factor 8 (Data, Observation & Summary) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 28, item 2 = item 29, item 3 = item 30 and item 4 = item 31.

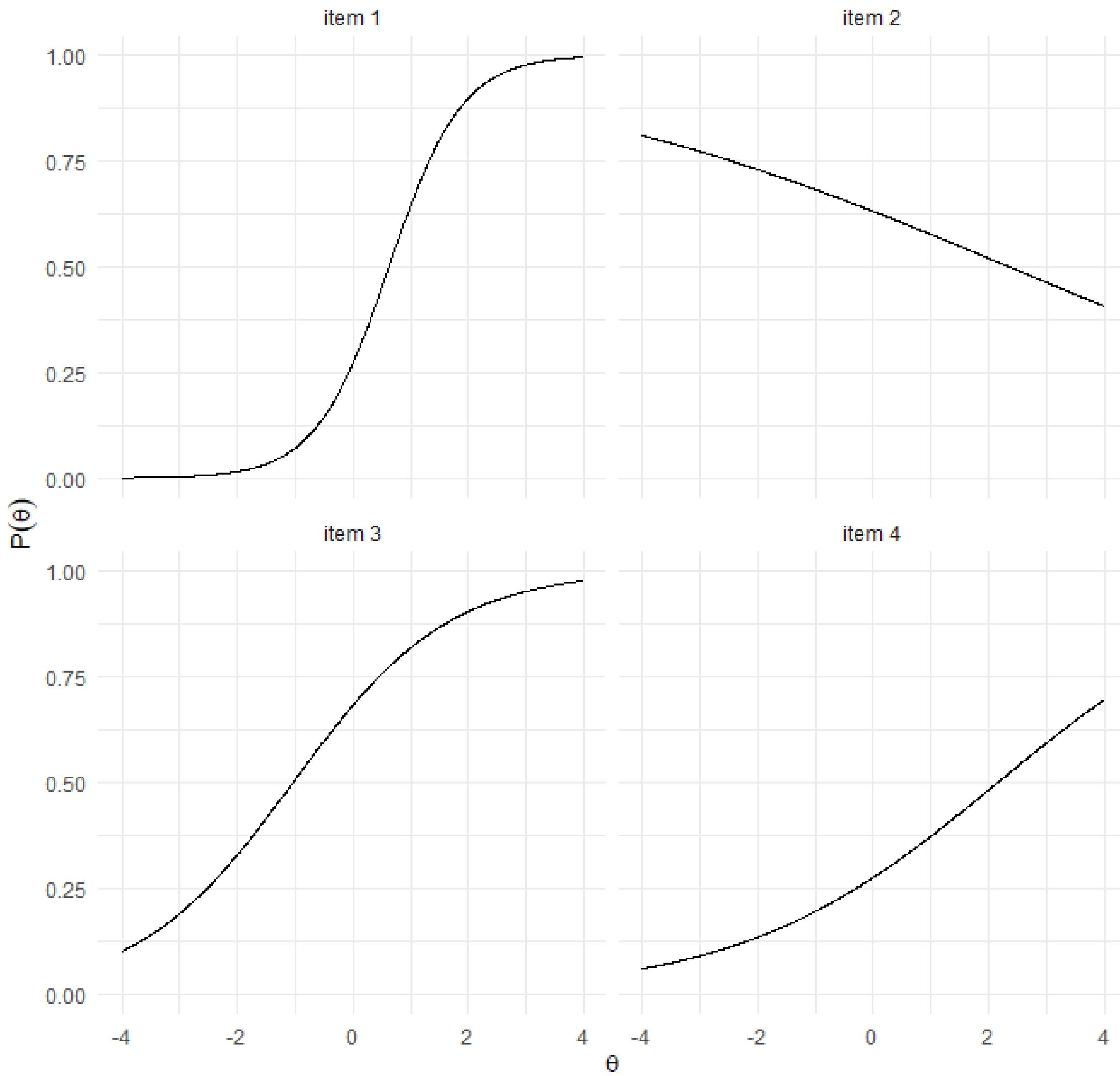


Figure B9

Item Characteristics Curves of items 32 to 34 in Factor 9 (Asymmetry) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item32, item 2 = item 33 and item 3 = item 34.

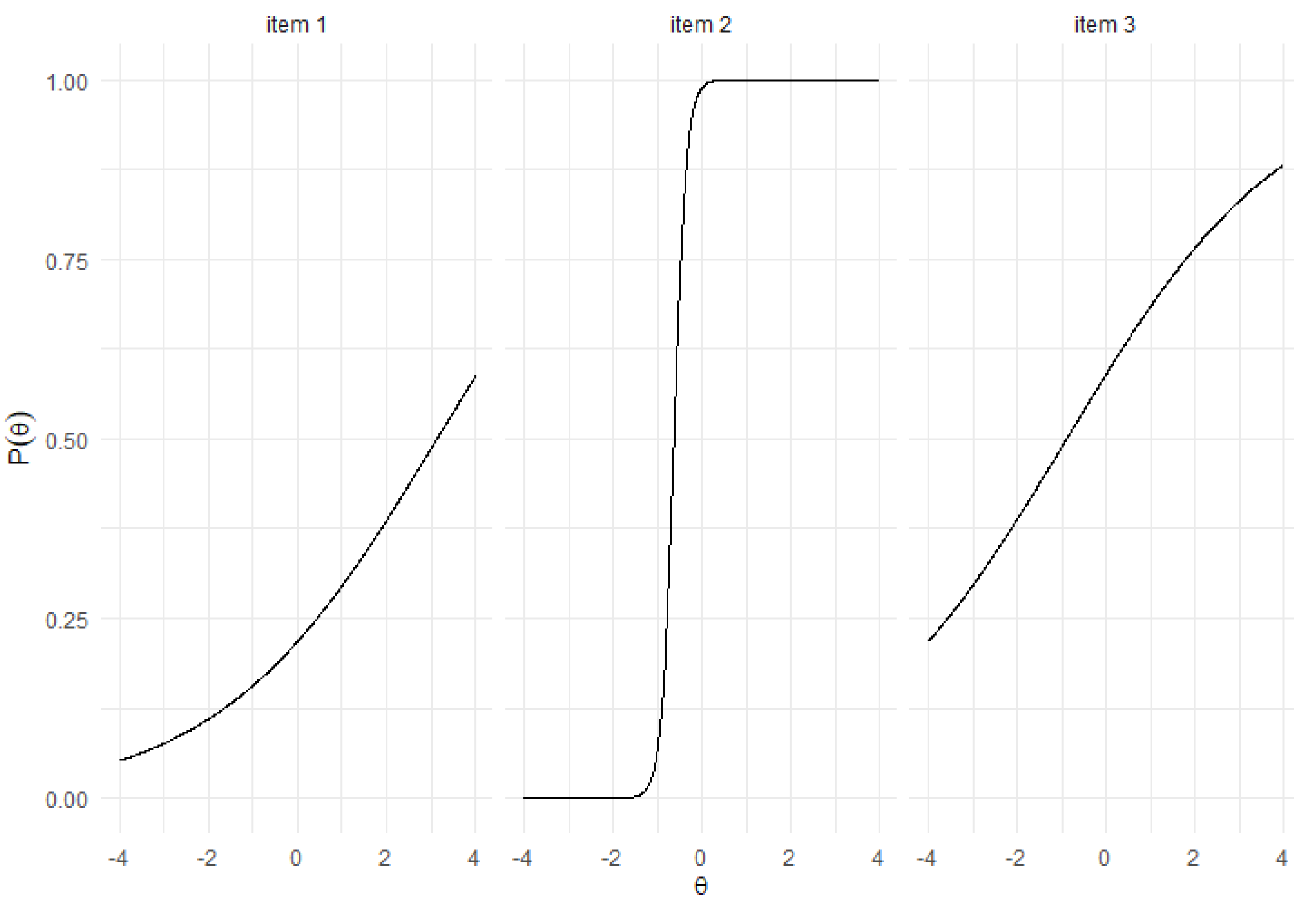


Figure B10

Item Characteristics Curves of items 35 to 37 in Factor 10 (Variability & Dispersion) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 35, item 2 = item 36 and item 3 = item 37.

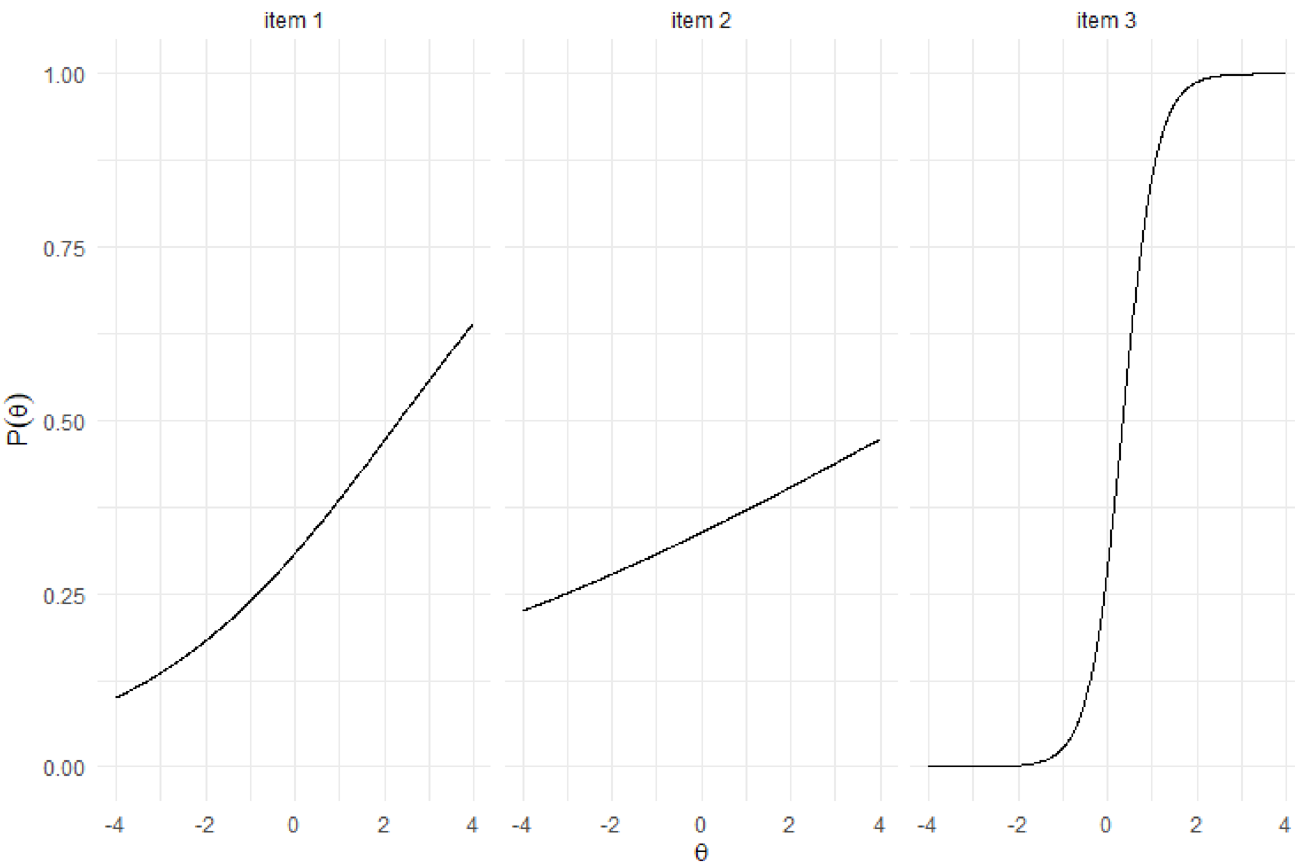


Figure B11

Item Characteristics Curves of items 40 to 43 in Factor 12 (Shared variance & Correlation) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 40, item 2 = item 41, item 3 = item 42 and item 4 = item 43.

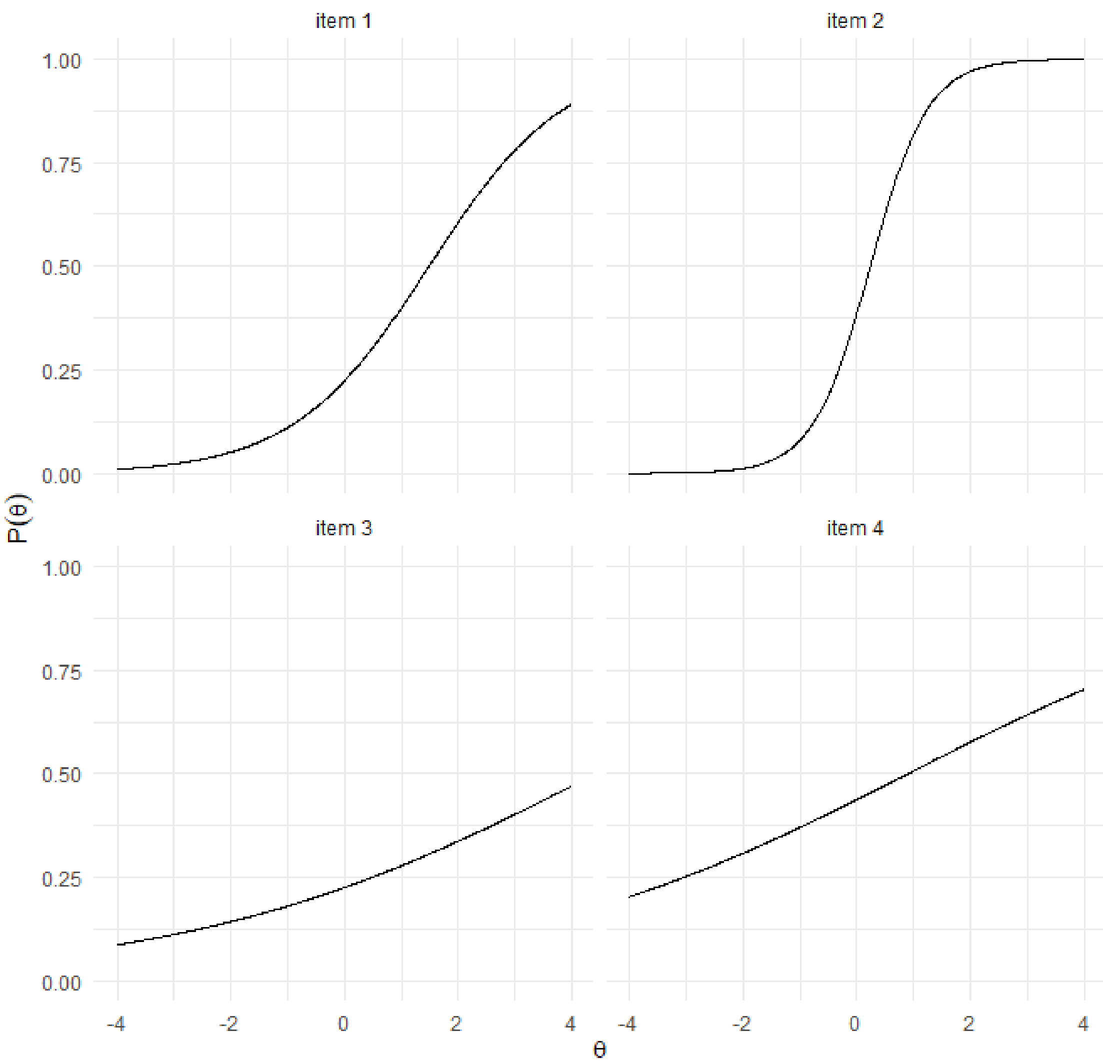


Figure B12

Item Characteristics Curves of items 44 to 46 in Factor 13 (Central tendency) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 44, item 2 = item 45 and item 3 = item 46.

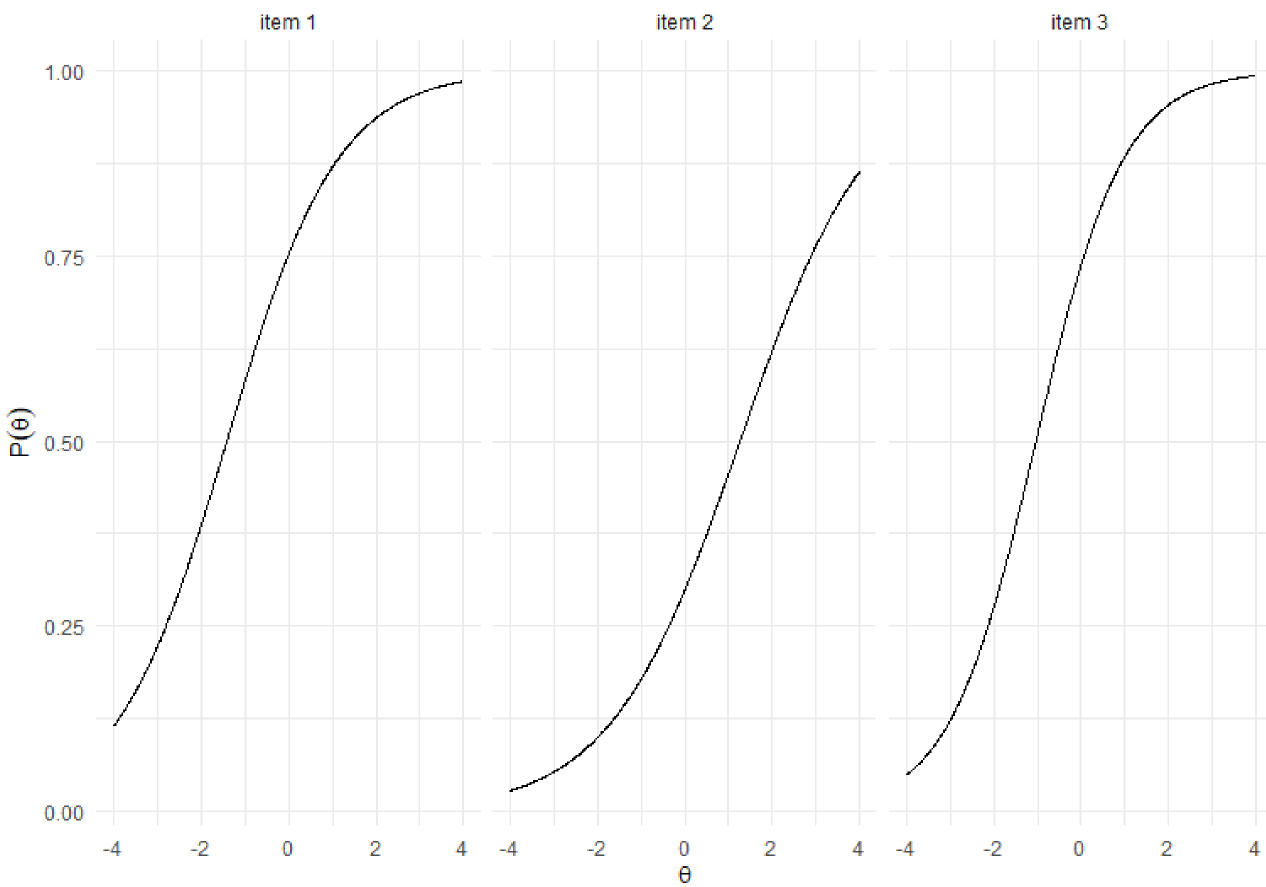


Figure B13

Item Characteristics Curves of items 47 to 49 in Factor 14 (Statistical Inference) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 47, item 2 = item 48 and item 3 = item 49.

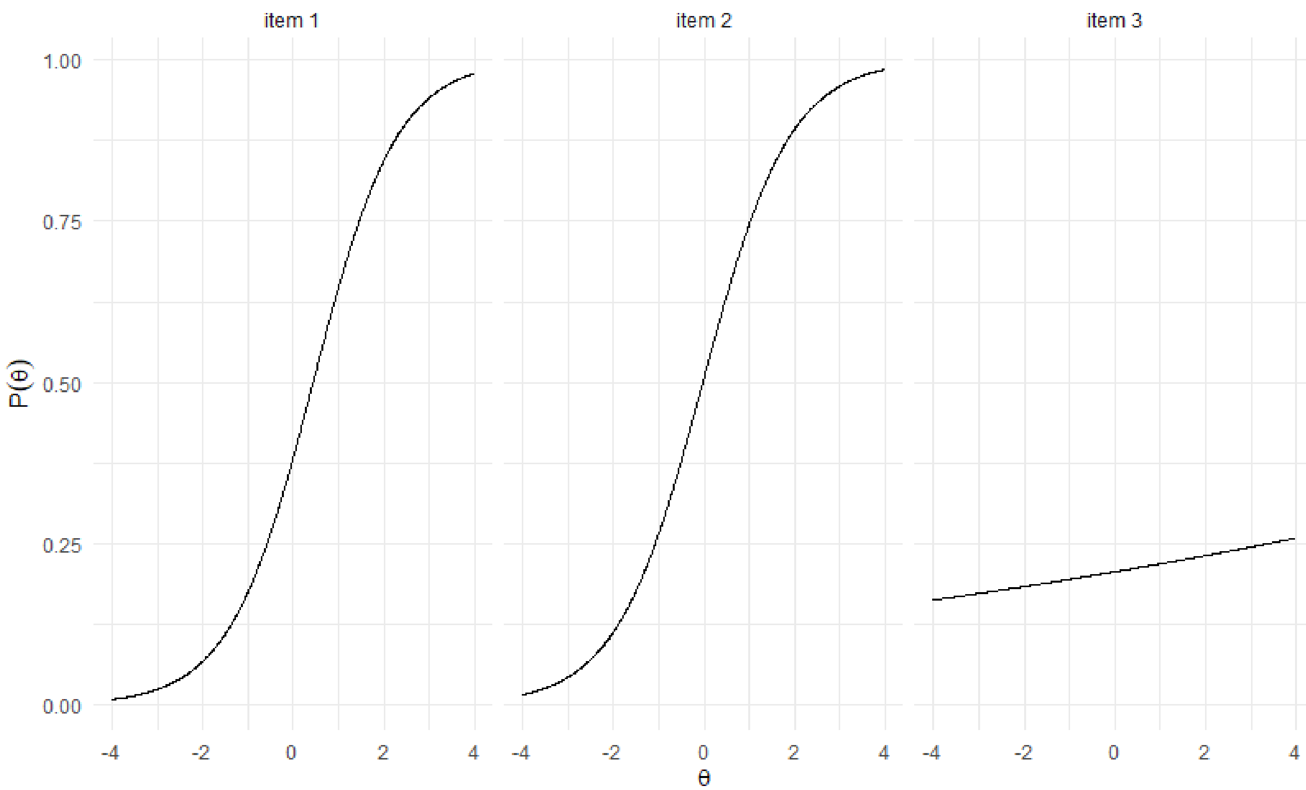


Figure B14

Item Characteristics Curves of items 50 to 52 in Factor 15 (NHST & p-value) of Iteration 1. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 50, item 2 = item 51 and item 3 = item 52.

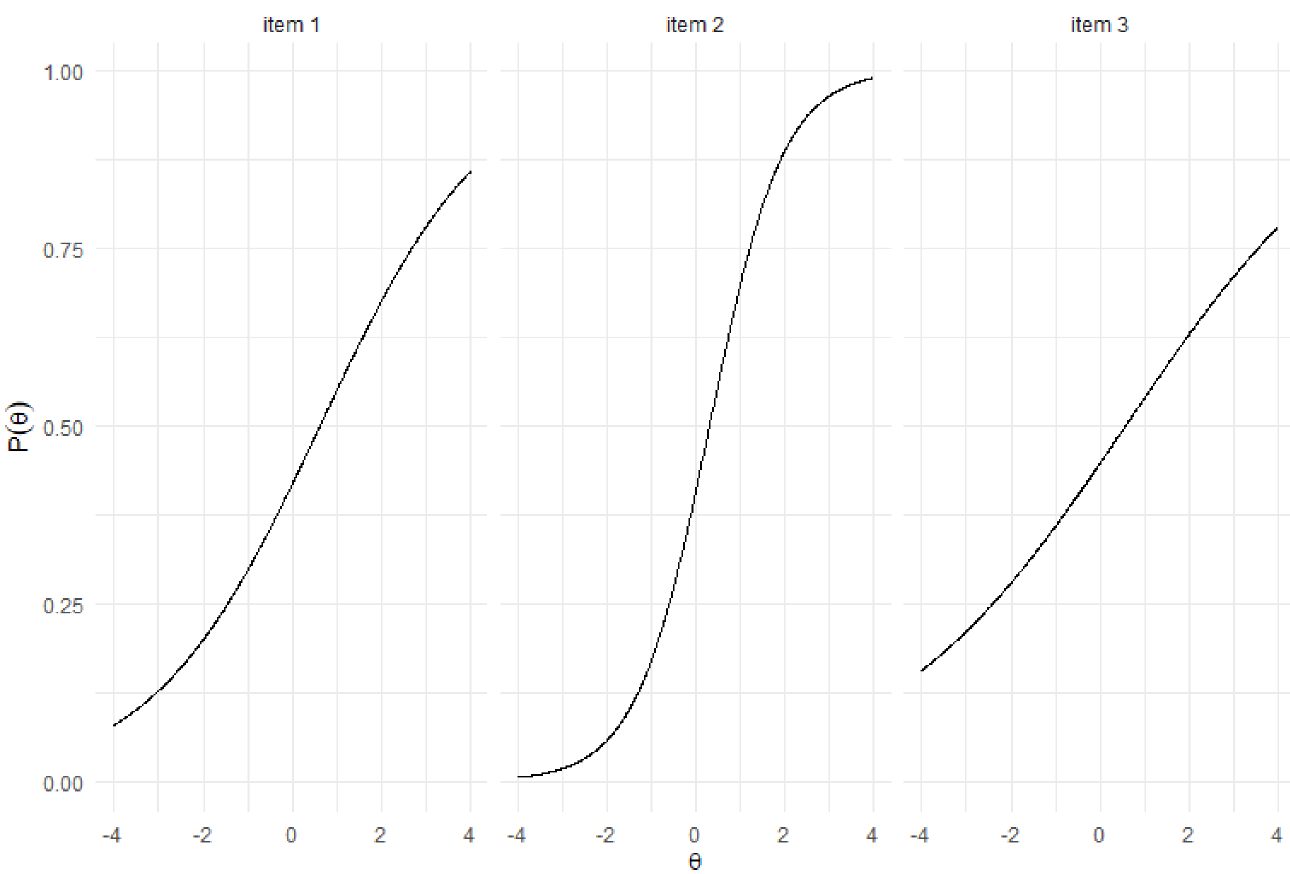


Figure B15

Item Characteristics Curves of items 1 to 4 in Factor 1 (Population) of Iteration 2. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability.

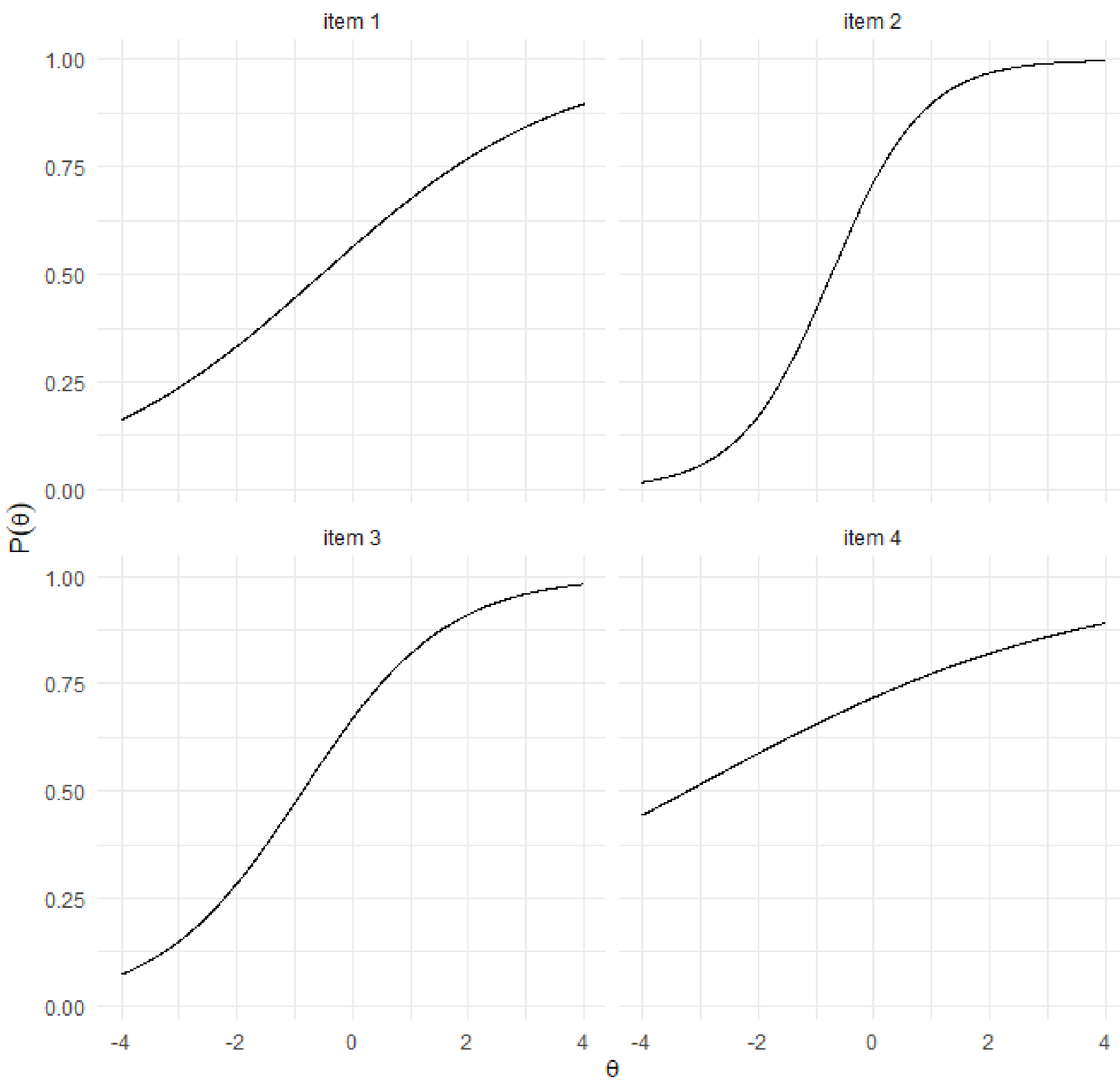


Figure B16

Item Characteristics Curves of items 5 to 8 in Factor 2 (Randomness & Probability) of Iteration 2. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 5, item 2 = item 6, item 3 = item 7 and item 4 = item 8.

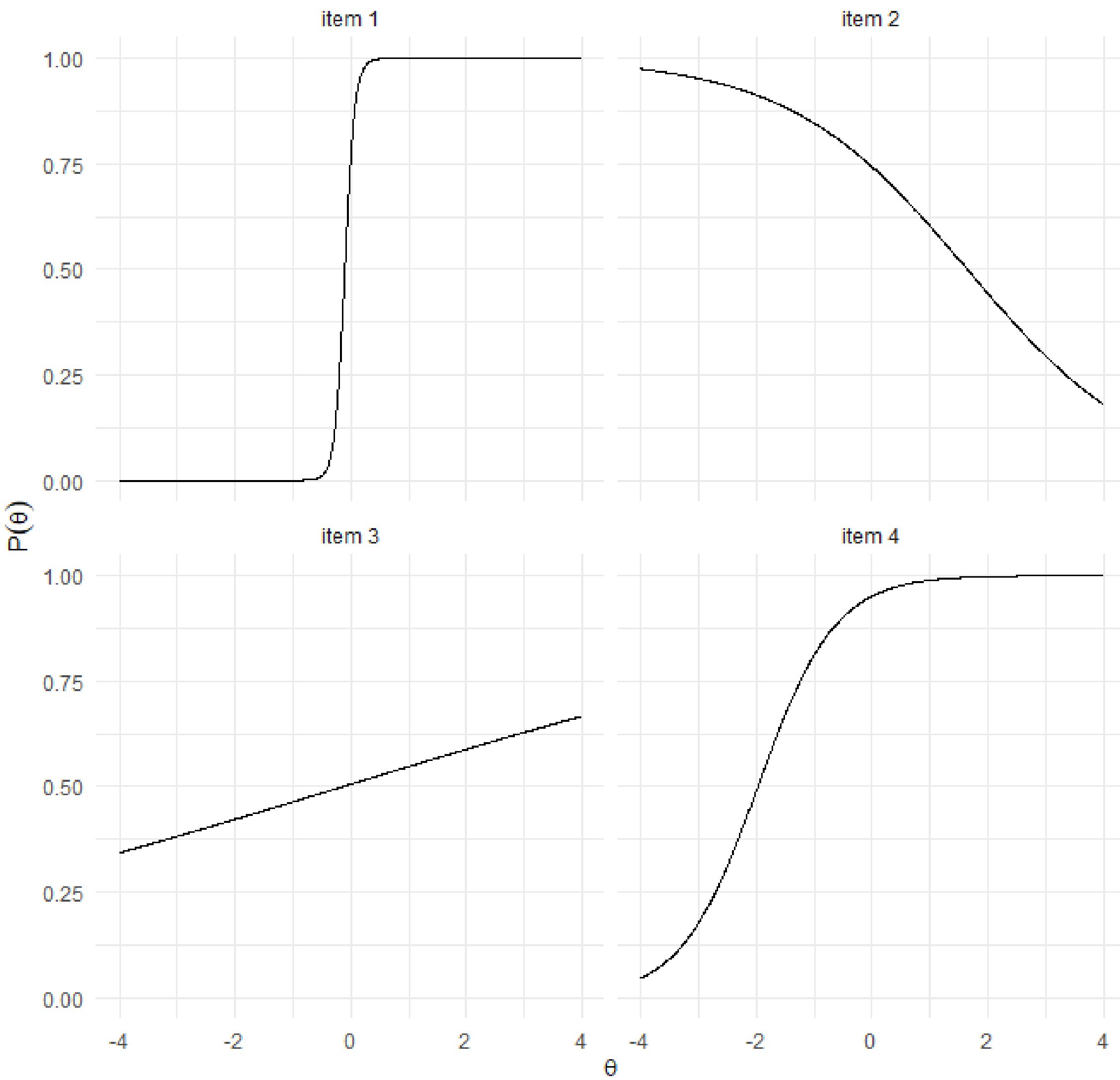


Figure B17

Item Characteristics Curves of items 9 to 11 in Factor 3 (Sample & Sampling) of Iteration 2. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 9, item 2 = item 10 and item 3 = item 11.

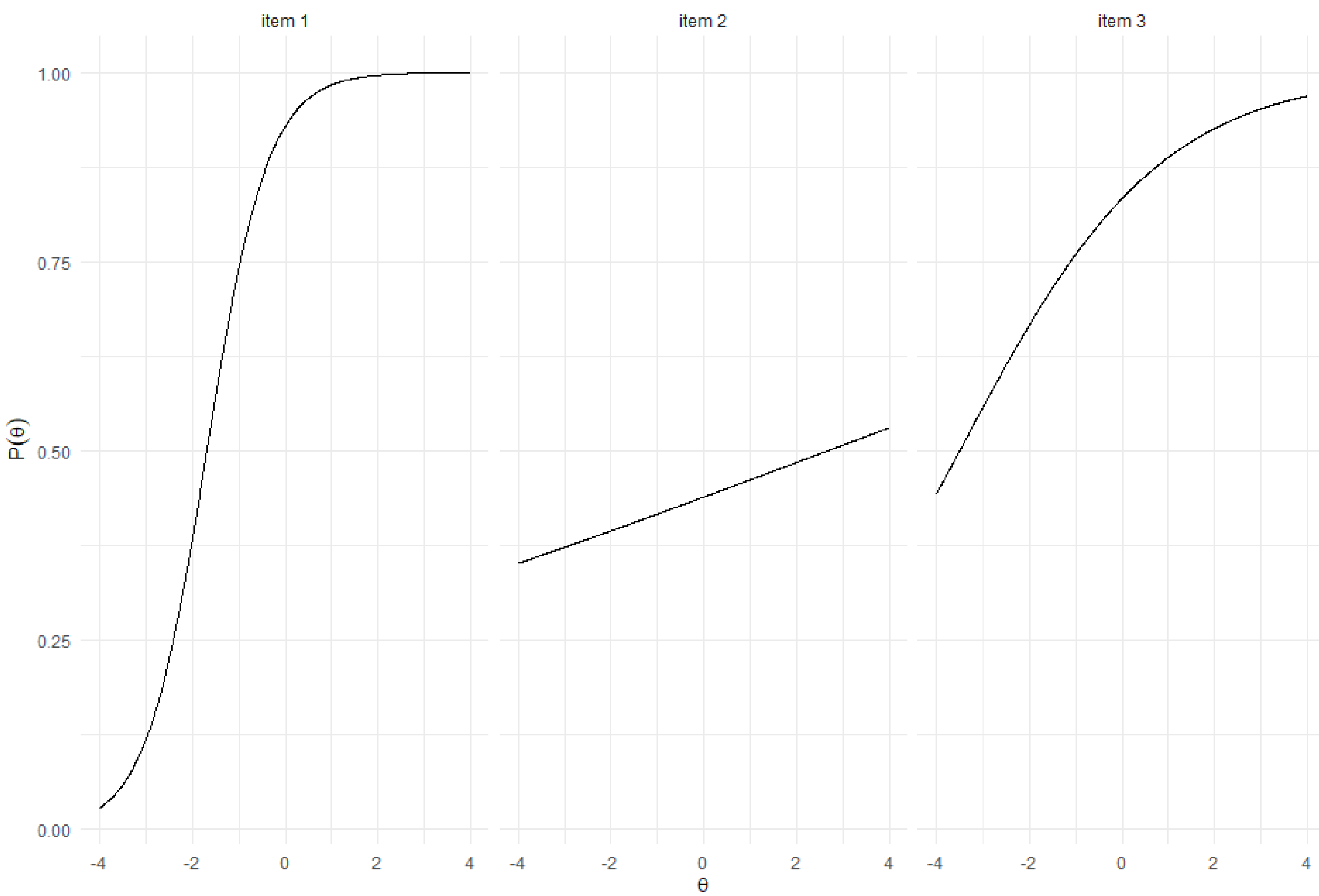


Figure B18

Item Characteristics Curves of items 12 to 14 in Factor 4 (Variability & Dispersion) of Iteration 2. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 12, item 2 = item 13 and item 3 = item 14.

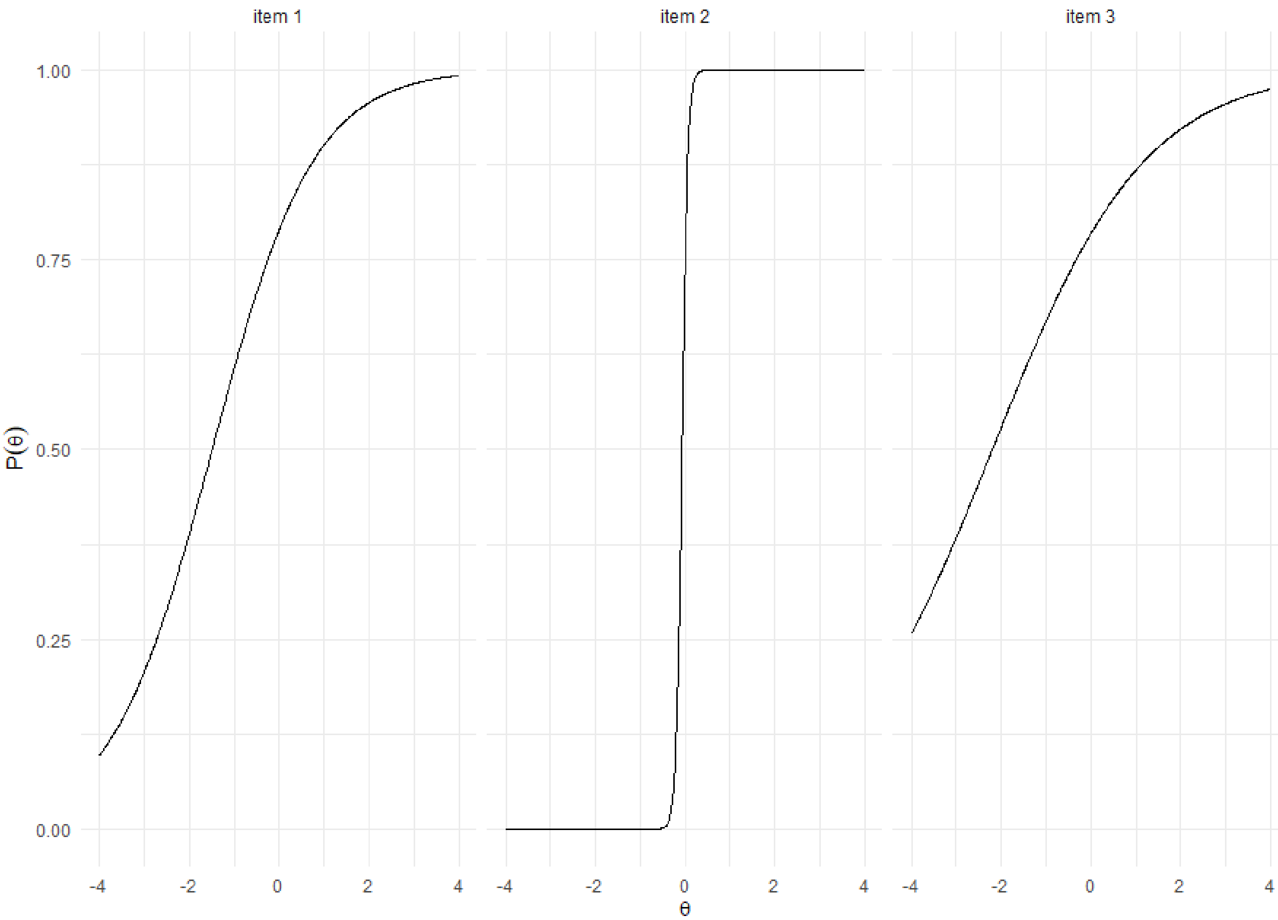


Figure B19

Item Characteristics Curves of items 15 to 17 in Factor 5 (Central tendency) of Iteration 2. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 15, item 2 = item 16 and item 3 = item 17.

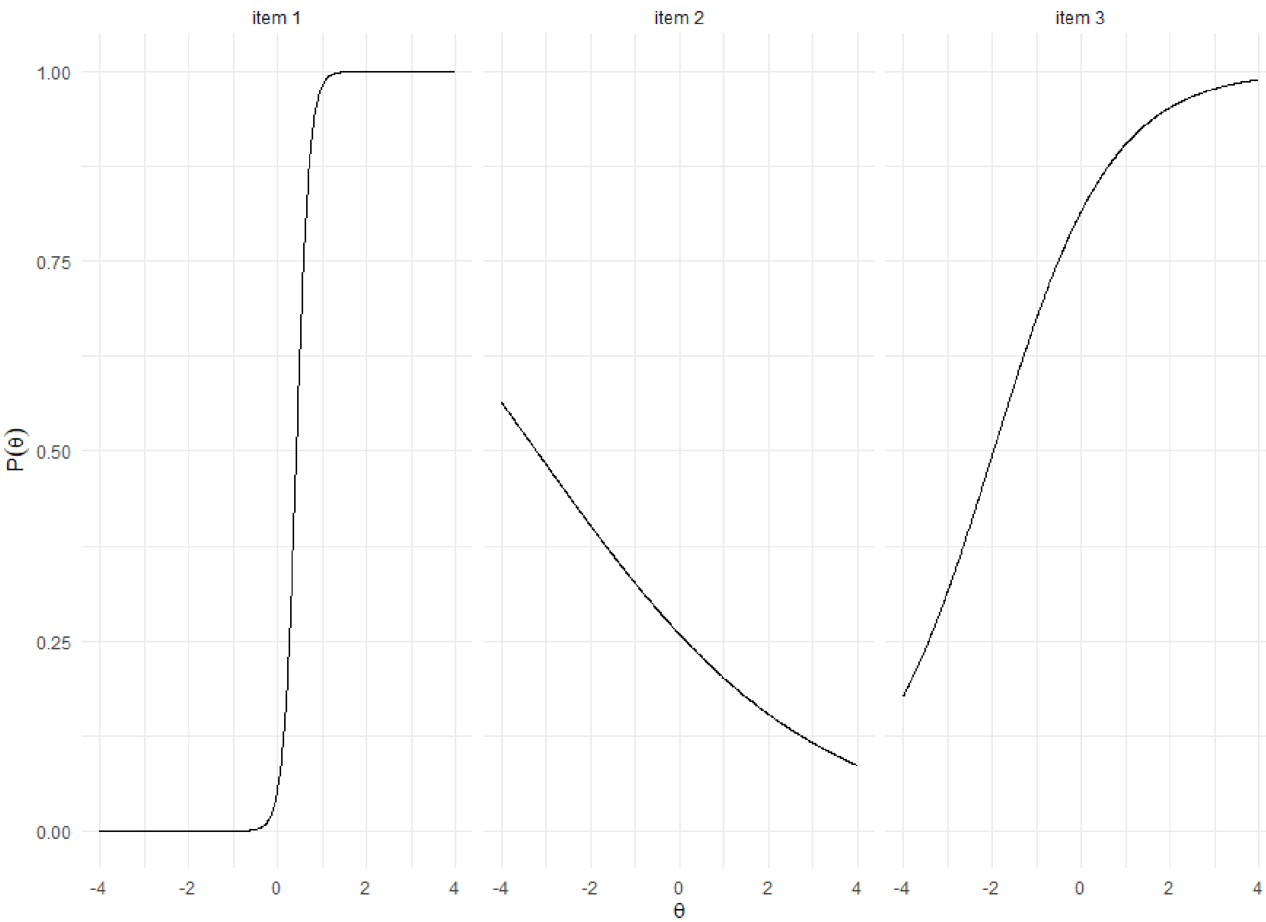


Figure B20

Item Characteristics Curves of items 1 to 12 in Factor 1 (Population & Sample) of Iteration 3. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. The items are in ascending order based on their first digit (for two-digit items).

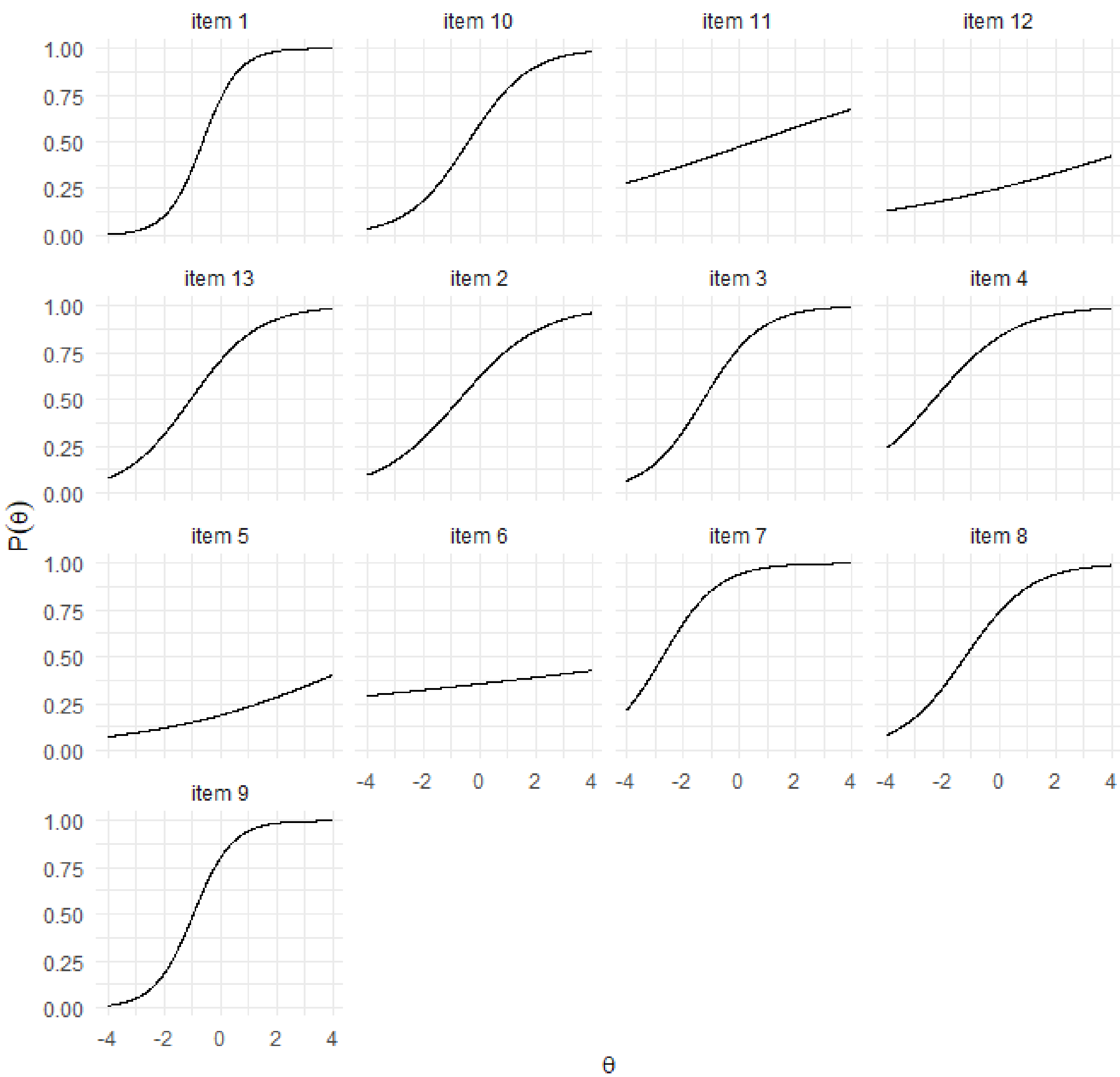


Figure B21

Item Characteristics Curves of items 13 to 21 in Factor 2 (Probability & Randomness) of Iteration 3. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 13, item 2 = item 14, item 3 = item 15, item 4 = item 16, item 5 = item 17, item 6 = item 18, item 7 = item 19, item 8 = item 20, item 9 = item 21.

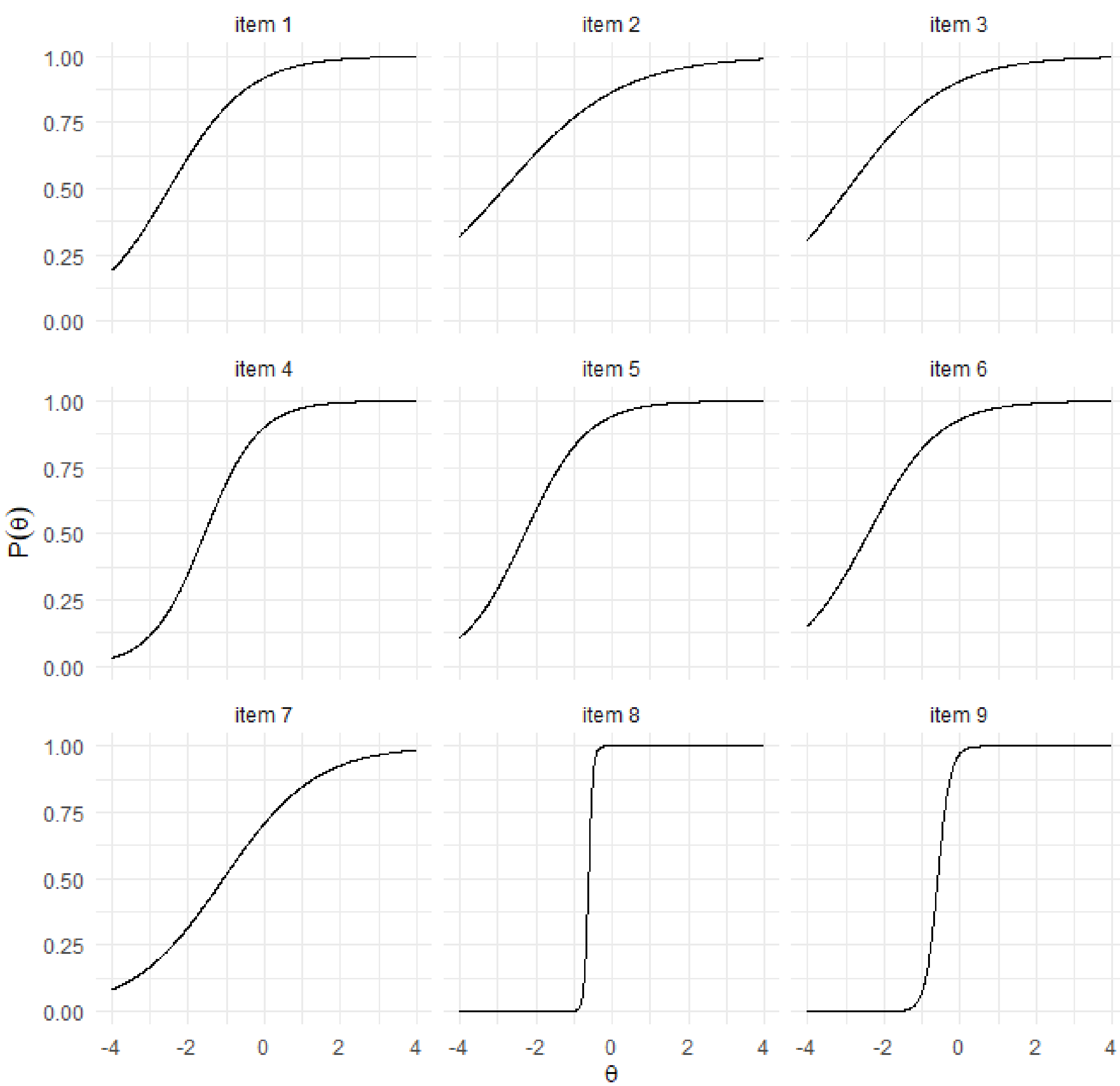


Figure B22

Item Characteristics Curves of items 22 to 27 in Factor 3 (Variability & Dispersion) of Iteration 3. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 22, item 2 = item 23, item 3 = item 24, item 4 = item 25, item 5 = item 26, item 6 = item 27.

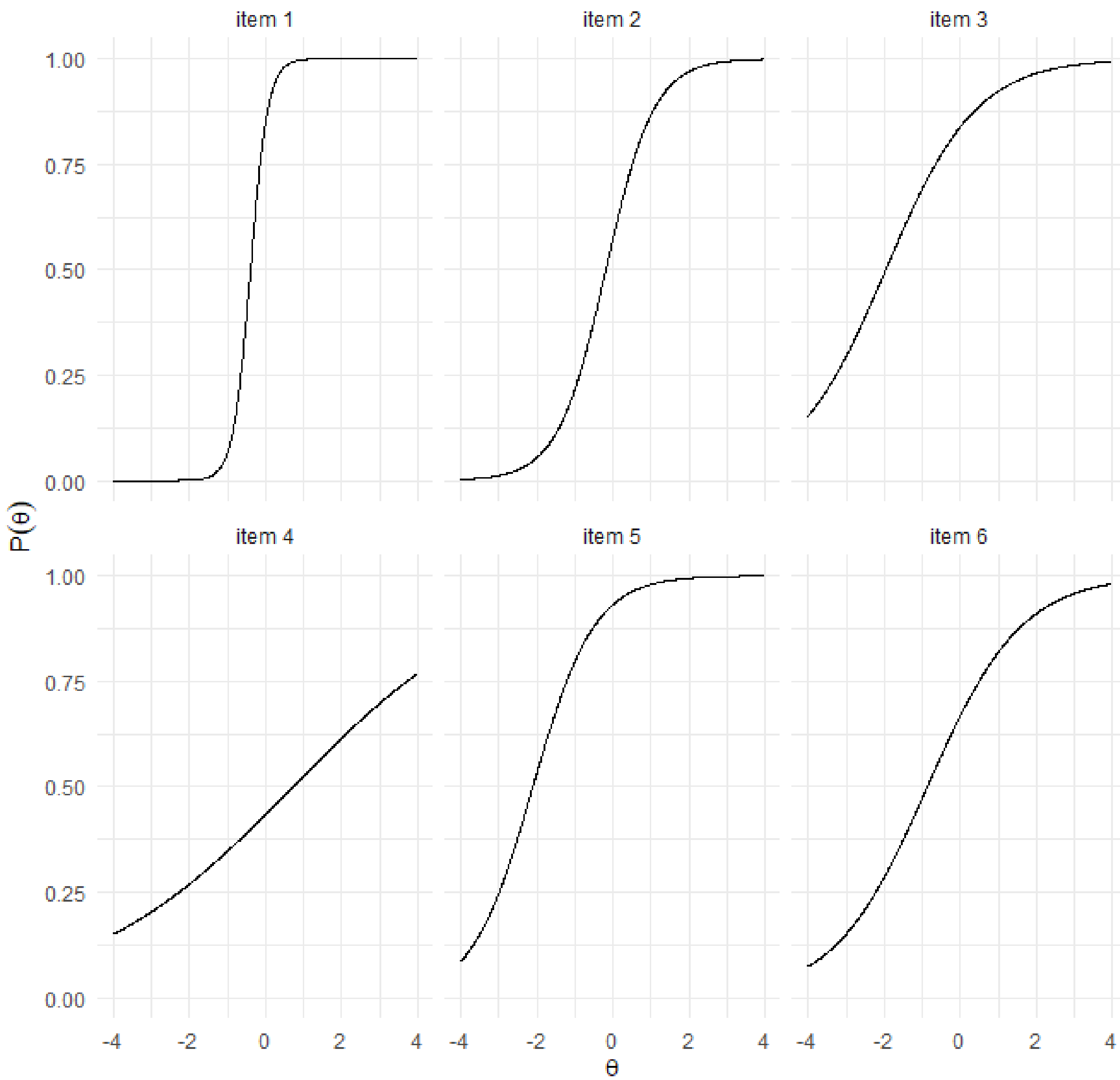


Figure B23

Item Characteristics Curves of items 28 to 33 in Factor 4 (Central tendency) of Iteration 3. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 28, item 2 = item 29, item 3 = item 30, item 4 = item 31, item 5 = item 32, item 6 = item 33.

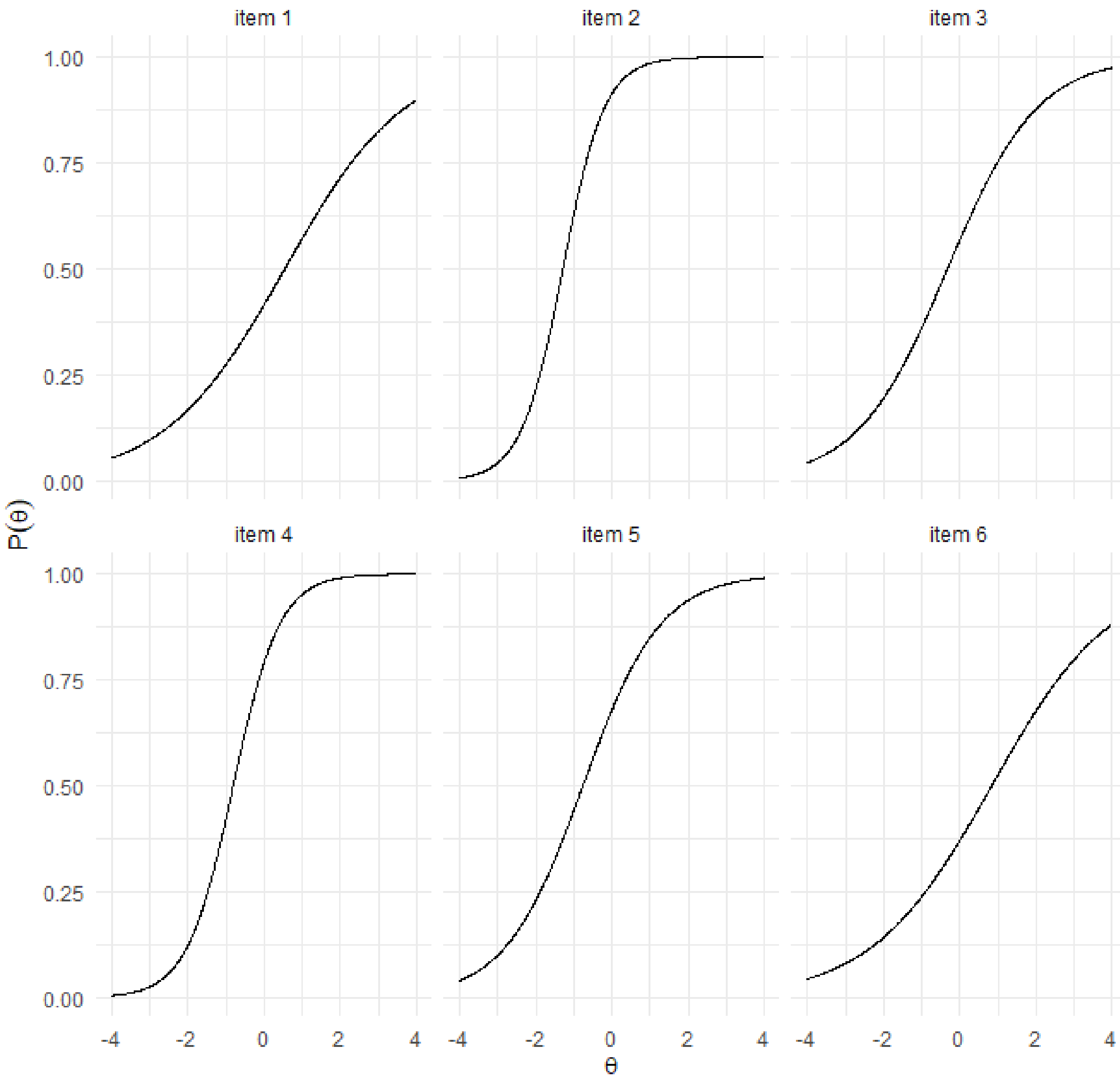


Figure B24

Item Characteristics Curves of items 1 to 12 in Factor 1 (Population & Sample) of Iteration 4. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. The items are in ascending order based on their first digit (for two-digit items).

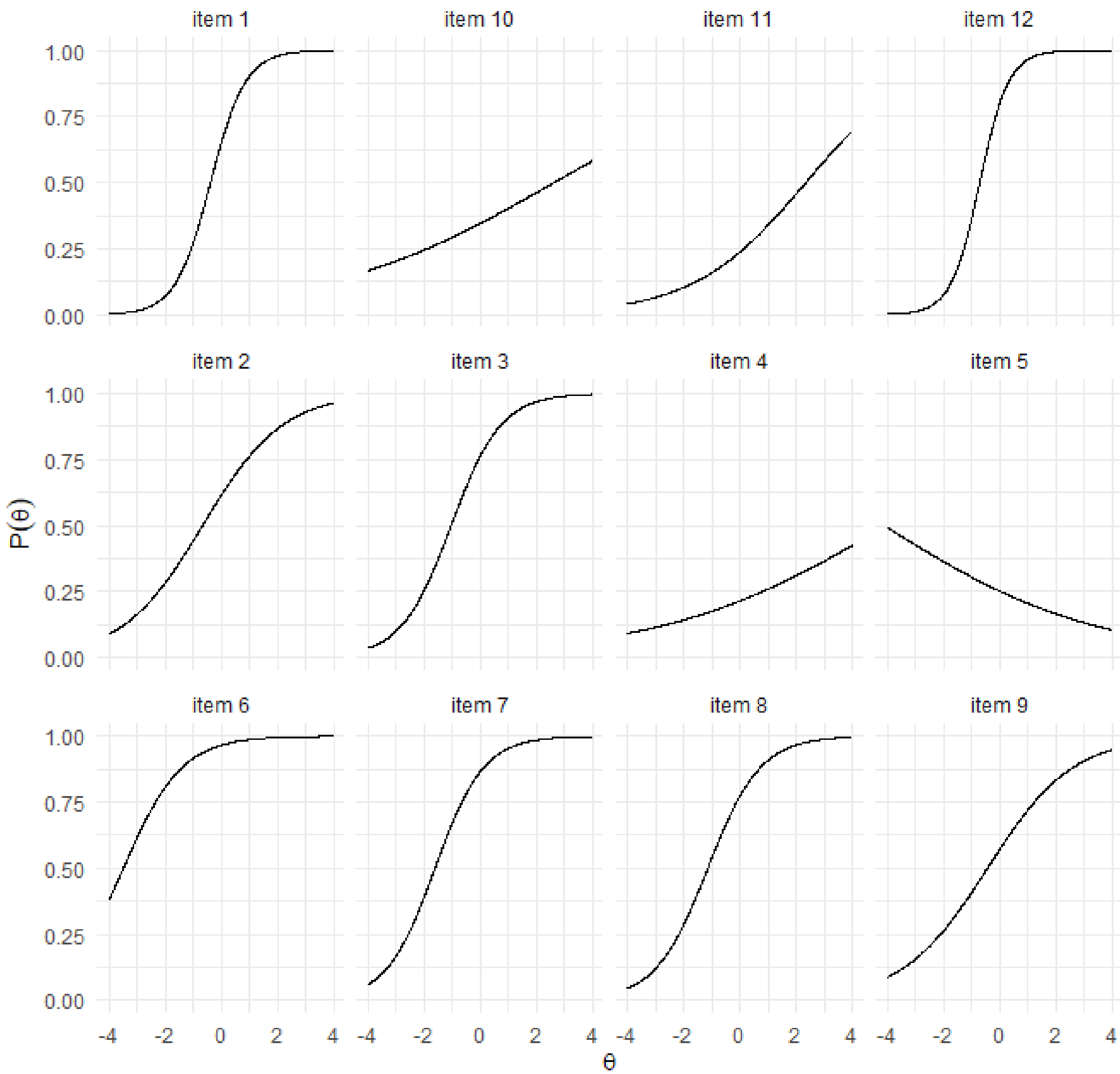


Figure B25

Item Characteristics Curves of items 13 to 21 in Factor 2 (Randomness & Probability) of Iteration 4. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 13, item 2 = item 14, item 3 = item 15, item 4 = item 16, item 5 = item 17, item 6 = item 18, item 7 = item 19, item 8 = item 20 and item 9 = item 21.

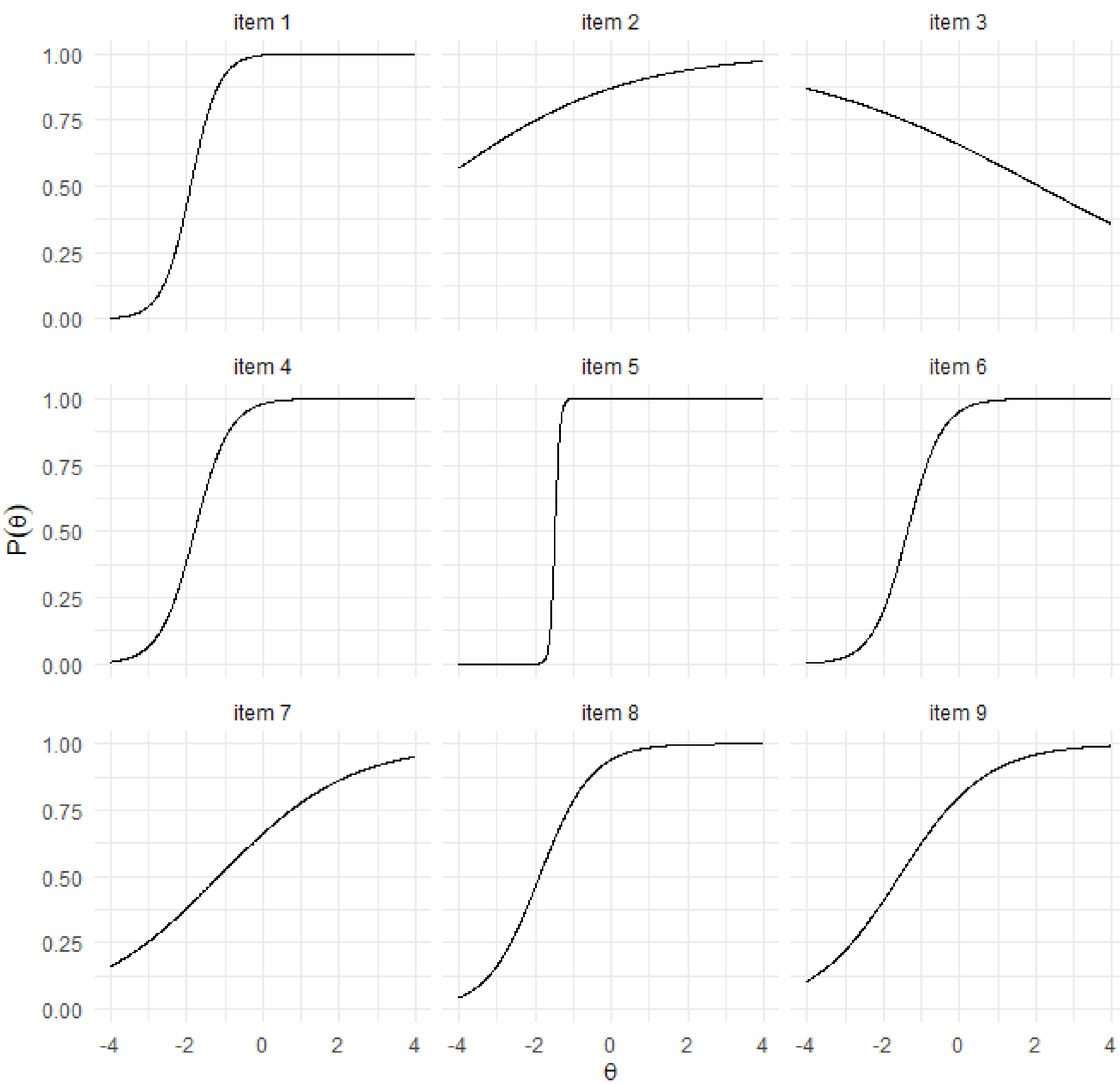


Figure B26

Item Characteristics Curves of items 22 to 27 in Factor 3 (Variability & Dispersion) of Iteration 4. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 22, item 2 = item 23, item 3 = item 24, item 4 = item 25, item 5 = item 26 and item 6 = item 27.

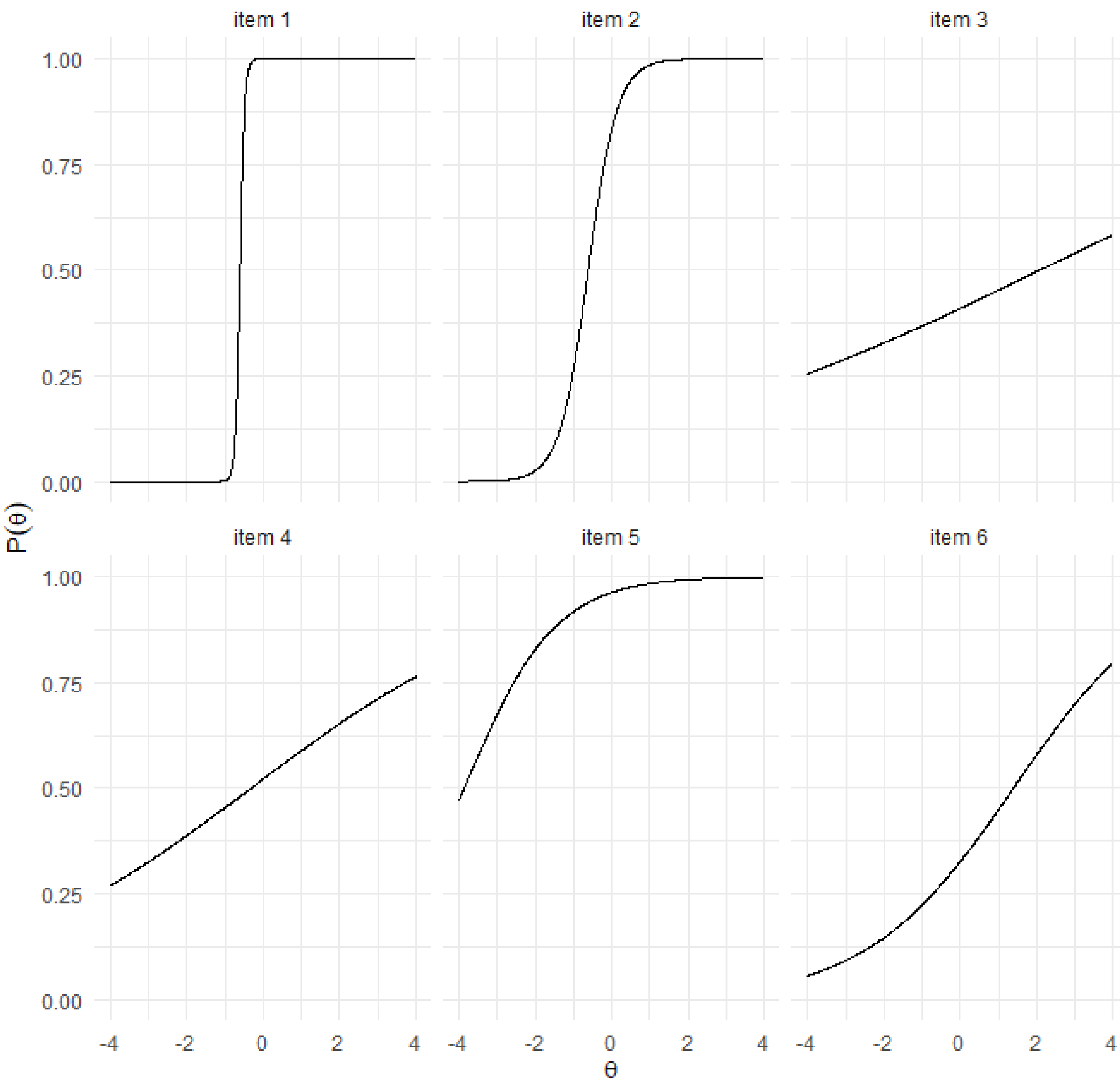


Figure B27

Item Characteristics Curves of items 28 to 33 in Factor 4 (Central tendency) of Iteration 4. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 28, item 2 = item 29, item 3 = item 30, item 4 = item 31, item 5 = item 32, item 6 = item 33.

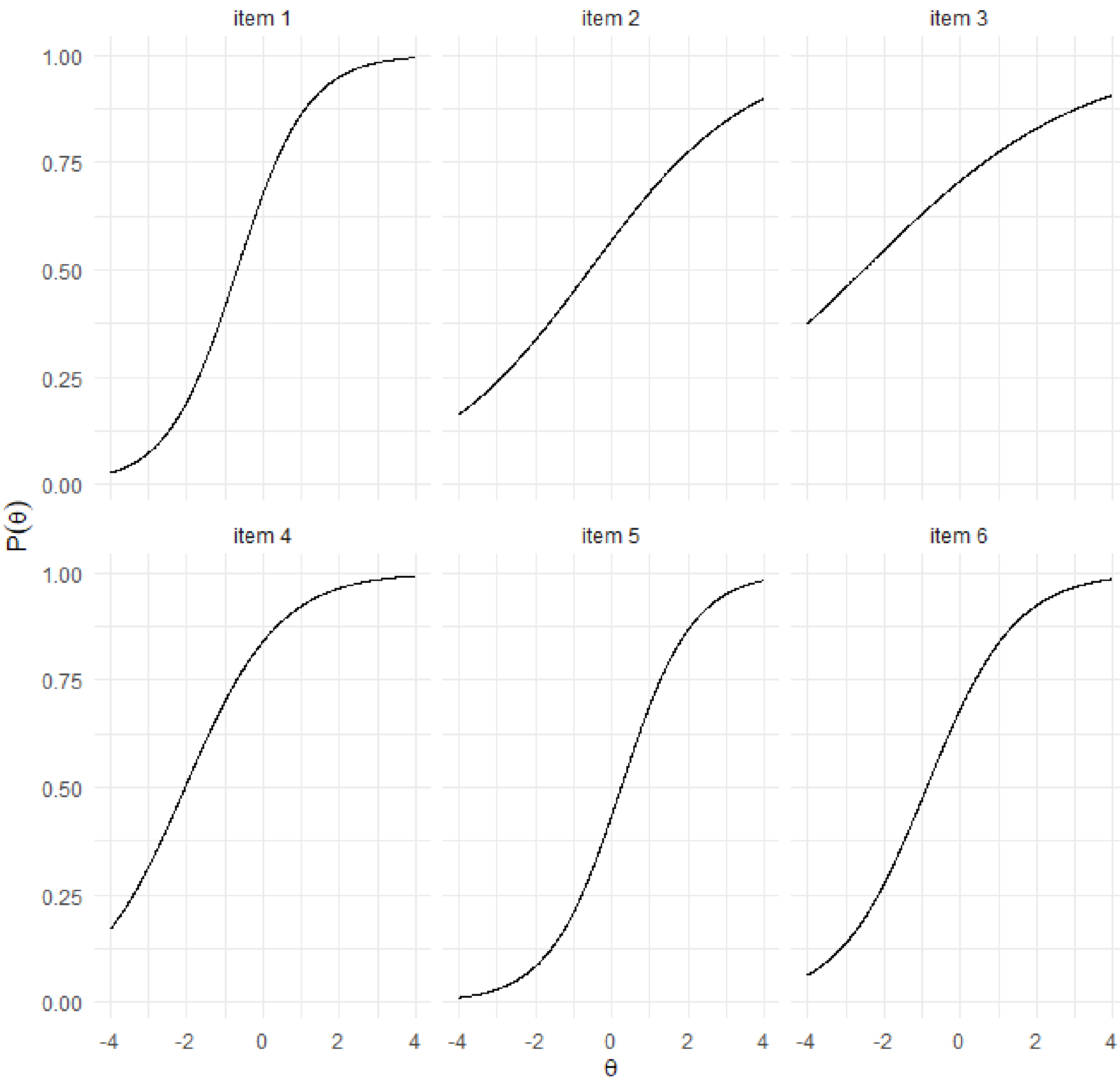


Figure B28

Item Characteristics Curves of items 1 to 9 in Factor 1 (Population & Sample) of Iteration5. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 1, item 2 = item 2, item 3 = item 3, item 4 = item 4, item 5 = item 5, Item 6 = item 6, item 7 = item 7, item 8 = item 8, item 9 = item 9.

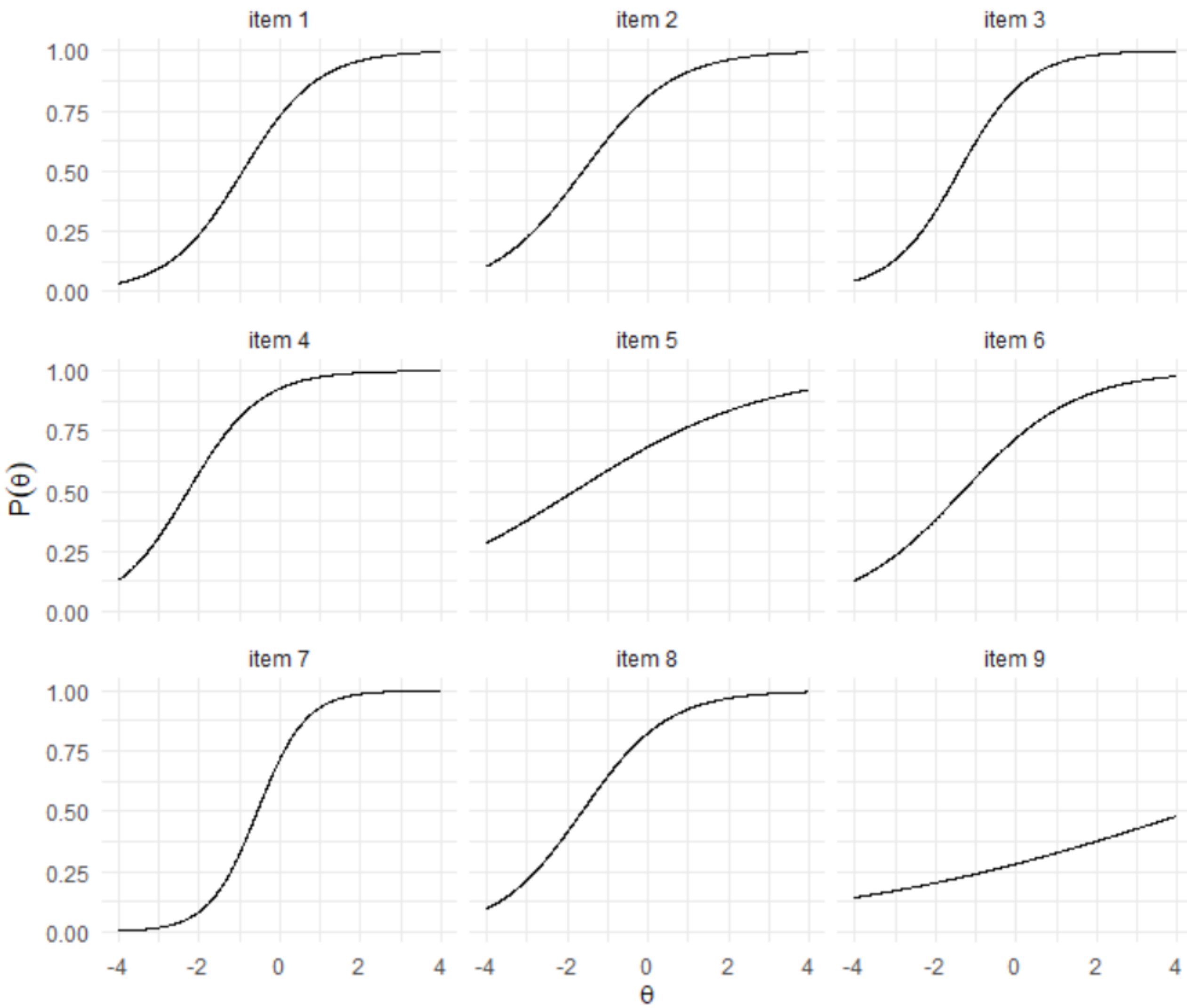


Figure B29

Item Characteristics Curves of items 10 to 13 in Factor 2 (Randomness & Probability) of Iteration 5. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 10, item 2 = item 11, item 3 = item 12, item 4 = item 13.

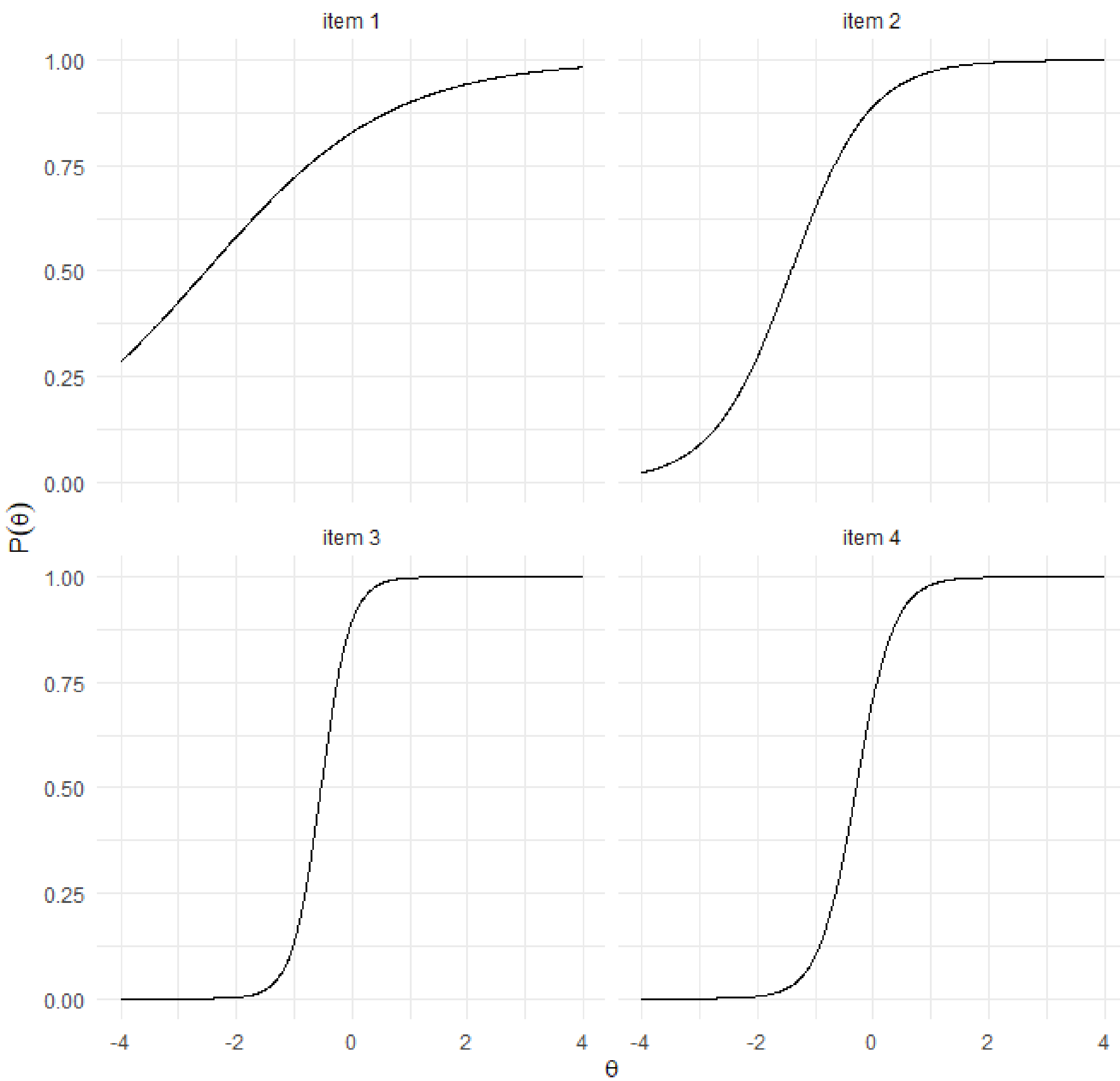


Figure B30

Item Characteristics Curves of items 14 to 18 in Factor 3 (Variability & Dispersion) of Iteration 5. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 14, item 2 = item 15, item 3 = item 16, item 4 = item 17, item 5 = item 18.

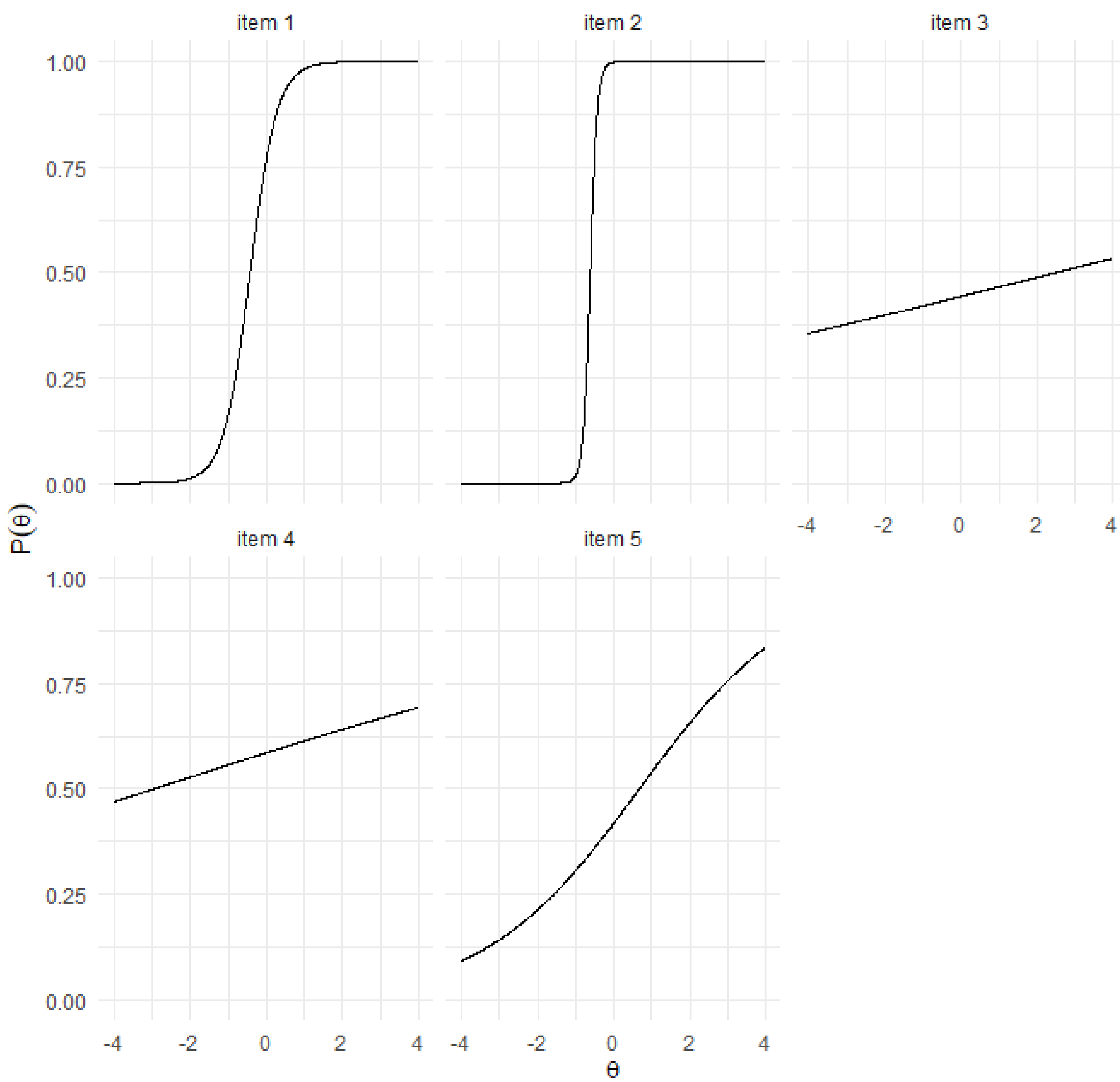


Figure B31

Item Characteristics Curves of items 19 to 22 in Factor 4 (Central tendency) of Iteration 5. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 19, item 2 = item 20, item 3 = item 21, item 4 = item 22.

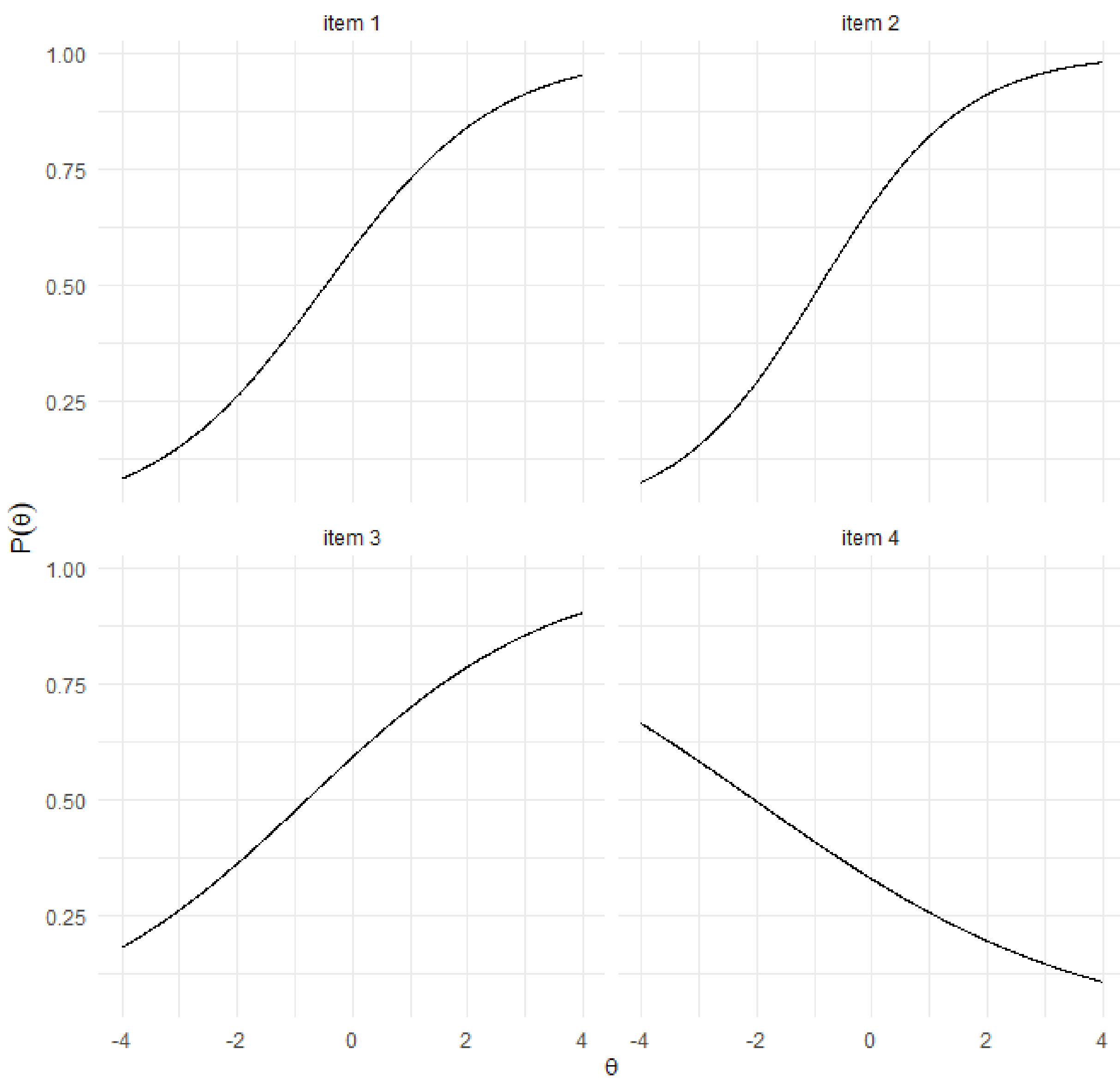


Figure B32

Item Characteristics Curves of items 23 to 26 in Factor 5 (Shared Variance & Correlation) of Iteration 5. The X axis shows ability levels, where 0 represents an average-ability person. The Y axis shows the success probability. Item 1 = item 23, item 2 = item 24, item 3 = item 25, item 4 = item 26.