Self-Serving Biases in Students' Evaluations of Teaching: Examining the Impact of Self-Reported Narcissism and Shyness

Melissa Leanne Staddon, H. B.A.

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the Master of Arts in Education degree

Faculty of Education

University of Ottawa

Acknowledgements

First and foremost, I would like to extend my sincere thanks to my thesis supervisor, Dr. Tracy Vaillancourt for her support and belief in me since day one. I want to thank you for providing me with the opportunity to work and learn in the Brain and Behaviour Lab. I also thank my committee members, Dr. David Smith and Dr. Jessica Whitley, for your thoughtful and insightful feedback throughout every step of the thesis process. You invaluable expertise and time is greatly appreciated.

I would also like to thank my family, partner, friends, and colleagues for their continued faith and support in me over the years. You have been there for me throughout this entire process and believed from the start that I could do it. The road has been long, and it hasn't always been easy, but you have all stuck by my side right from the beginning and I cannot thank you enough for that.

Last, but not least, a sincere thank you to all of my friends and colleagues in the Brain and Behaviour Lab: Heather Brittain, Amanda Krygsman, Marielle Asseraf, Christine Blain-Arcaro, Eric Duku, Jennifer Hepditch, Irene Vitoroulis, Katie Dubeau, Weijun Wang, Carleigh Sanderson, J.D. Haltigan, and Courtney Briggs-Jude. It was a wonderful experience to get to know each and every one of you over the years. Thank you for your friendship, your support, your feedback, and your words of encouragement over the past two years.

Abstract

Students' evaluations of teaching (SETs) have been used as a primary means of evaluating the teaching quality at postsecondary institutions for a number of years. The research in this area has been consistently expanding in an effort to validate its use and efficacy. To date, there has been little consensus regarding the validity of these evaluations, especially with regards to the possible impact of extraneous factors. Researchers have previously found evidence of a link between the self-serving bias and grade expectations. The present study used an experimental design to examine the validity of SETs in the context of the self-serving bias. According to the self-serving bias, individuals will be more likely to attribute success internally but attribute failure externally. Specifically, the present study examined whether there were differences in the presence of the self-serving bias in relation to self-reported narcissism and shyness. Students at a large university aged 17 - 46 (N = 563) were asked to write a short essay on euthanasia and were randomly assigned to one of two conditions. In the first condition, students were assigned a low grade; in the second condition, they were assigned a high grade. When they were given their essay grade back, students had an opportunity to rate their evaluator. Results indicated that students were more likely to externalize their essay grade when in the negative condition whereby they attributed their grade to reasons outside their control versus the positive condition, irrespective of their level of self-reported narcissism or shyness. These results suggest that the self-serving bias does exist within SETs and calls into question the validity of these evaluations. The findings from this study highlight the need for further research into the role that student characteristics play with regards to SETs.

Self-Serving Biases in Students' Evaluations of Teaching: Examining the Impact of Self-Reported Narcissism and Shyness

Quality assurance at the level of postsecondary education has been an important and debated topic for decades (Tam, 2001; Woodhouse, 1998). Much has been written about issues of accountability within the realm of academia, including institutional and program accreditation (Association of Universities and Colleges of Canada (AUCC), 2012) and public funding of higher education (Simpson & Siguaw, 2000). There is no doubt that the areas within postsecondary institutions (e.g., quality of teaching, external funding, student and employee satisfaction, etc.), for which the public and government have demanded accountability, constitute a lengthy list. Governmental agencies and the public at large are holding postsecondary institutions accountable to measurable outcomes (Kelly, 2012; Simpson & Siguaw, 2000) and the methods used to measure and evaluate the effectiveness of postsecondary institutions and their programs has received much attention in the literature.

One of the areas of quality assurance that has consistently caught the attention of researchers and stakeholders is the quality of teaching at universities and colleges. The measurement of teaching effectiveness is particularly relevant given the recent increase in focus on teaching and learning in universities, especially those in North America (Kelly, 2012). The predominant means for examining and assessing instructor performance has been student evaluations of teaching (SETs). SETs were first adopted by a small number of postsecondary institutions in the United States during the mid-1920s as a way to measure instructional effectiveness (d'Apollonia & Abrami, 1997; Doyle, 1983). Throughout the late 1960s and early 1970s, there was a dramatic increase in the number of research articles published on the topic of SETs, which extensively examined the psychometric properties, usefulness, and validity of these

measures. Areas that were commonly examined in these evaluations include, but are not limited to, workload of the course, clarity of instruction, approachability of instructor, organization of course material, and fairness of formal evaluation (Cashin, 1999; Marsh & Roche, 1997).

The application of SETs has now become nearly universal at North American postsecondary institutions and student surveys of instruction have become the most widely used method of assessing teaching and instructor effectiveness (Centra, 2003; Kelly, 2012; Wolfer & Johnson, 2003). Research in both Canada and the United States has found that at least 90% of postsecondary institutions use student rating questionnaires, either alone, or in conjunction with other methods to evaluate teaching (Saroyan & Amundsen, 2001; Seldin, 1998). The implications that result from the use of SETs are vast. Ideally, these evaluations are a valuable tool used to improve both student learning outcomes and teaching performance, but in many cases the data collected are used to make decisions related to future employment. These decisions include retention, tenure, and promotion of faculty (Gravestock & Gregor-Greenleaf, 2008; Kelly, 2012; Zabaleta, 2007), and some researchers have argued that the results generated from SETs are used to make important personnel decisions without further analysis or interpretation (McKeachie, 1997; Zabaleta, 2007).

Evaluating teaching performance and effectiveness through SETs at North American postsecondary institutions will likely remain an important component of assessing teaching and instructor effectiveness, and it is important to determine how valid and effective these evaluations are. Given that SETs are one of the most important and, in many cases, the only measure used to evaluate the teaching effectiveness of instructors at postsecondary institutions it is not surprising that their use has been widely debated and researched. A simple literature search in an educational research database such as ERIC elicited over 43,000 results (retrieved

April 15, 2013). Although there is a tremendous amount of research on this topic, little agreement has been reached in the research community regarding the validity of SETs (e.g., Greenwald & Gillmore, 1997; Marsh & Roche, 1997; Shevlin, Banyard, Davies, & Griffiths, 2000). Indeed, there remains the unresolved issue of whether SETs are an accurate reflection of true teaching effectiveness or whether they are contaminated by factors unrelated to teaching (Chacko, 1983). As such, the focus of the current study was to examine whether external factors, such as student personality characteristics, impact the outcome of written evaluations.

The Impacts of Extraneous Factors on Students' Evaluations of Teaching

The value of SETs is based on the notion that the evaluations students give their instructors is based solely on their teaching effectiveness and performance and not on external factors outside of an instructor's control. However, a large body of research has pointed to the fact that SETs can be, and often are, affected by factors not related to actual teaching effectiveness. These different factors include those that are teacher related (e.g., Felton, Mitchell, & Stinson, 2004; Freng &Webber, 2009; Kierstead, D'Agostino, & Dill, 1988; Sinclair & Kunda, 2000), course related (e.g., Addison, Best, & Warrington, 2006; Feldman, 1978; Gage, 1961; Lovel & Haner, 1955; Scherr & Scherr, 1990), and student related (Marsh & Roche, 1997; Patrick, 2011).

Teacher-Related Factors

Teacher-related factors that may have an effect on SETs include the instructor's perceived level of physical attractiveness (Felton, Mitchell, & Stinson, 2004; Freng & Webber, 2009), whereby physically attractive instructors are rated more favourably than their presumably less attractive colleagues. Felton, Mitchell, and Stinson (2004) examined the relationship between perceived quality of teaching, easiness, and sexiness of professors through a well-

known web-based student evaluation measure (RateMyProfessor.com). Using a sample size of over 3000 professors from 25 post secondary institutions in the United States, the researchers found that, when grouped into sexy and non-sexy professors, students gave the sexy-rated professors higher quality and easiness scores. More recently, Felton and colleagues (2008) examined these same variables using a larger and improved database from the same web-based student evaluation measure. Consistent with their earlier work, results showed even stronger relations where "hot" professors were given higher quality and easiness scores than their less attractive colleagues. As can be clearly observed in the studies described above, the level of perceived physical attractiveness of the instructor can play a role in how they are evaluated by their students. Indeed, some research has shown that the attractiveness of an instructor is positively correlated with factors such as immediacy, approachability, likeability, and teaching effectiveness, as reported by students (Feeley, 2002; Gurung & Vespia, 2007; Rocca & McCroskey, 1999).

Another teacher-related factor that has been found to adversely affect SETs is the gender of the instructor, but the research findings in this area are mixed. Whereas some authors have found evidence supporting higher evaluation for male professors versus their female colleagues (e.g., Kierstead, D'Agostino, & Dill, 1988; Sinclair & Kunda, 2000), others have found no such relationship (see reviews by Feldman, 1992, 1993). Additionally, some researchers (e.g., Centra & Gaubatz, 2000; Feldman, 1997) have found that students tend to rank instructors of their own gender slightly higher.

In one study exploring teacher-related factors that may impact SETs, Kierstead and colleagues (1988) examined whether the gender of an instructor had an impact on how they were evaluated by students. In addition, the authors also examined whether social contact between the

instructor and students and the instructor's facial expression (smiling versus neutral) had an impact on ratings. Overall, it was found that male instructors were judged as being more effective than female instructors. It is interesting to note, however, that behaviours indicative of friendliness (e.g., smiling) toward students elevated the ratings of female instructors but not male instructors.

Sinclair and Kunda (2000) examined whether ratings of instructors were influenced by gender and grades given to students. Using a sample of 570 course evaluations for 318 different instructors (230 men and 82 women), the authors investigated whether an instructor's ratings would be high if the participant had received a high grade in their class. Furthermore, the authors wanted to determine if this rating was influenced by the gender of the instructor. Consistent with previous research, Sinclair and Kunda found that participants viewed female instructors as less competent than their male colleagues after receiving low grades. Interestingly, a significant interaction was found between grade and gender of the instructor whereas when students received a high grade, the gender difference was no longer present.

Alternatively, two reviews examining the laboratory and experimental research on students' preconceptions of male and female college professors found that, in a majority of studies, global evaluations did not differ (Feldman, 1992, 1993). Feldman referred to global evaluations as the overall evaluation given to a professor. This finding held true for both male and female students, although students did tend to rate same-sex professors slightly higher than opposite-sex professors. Feldman did find, however, that in the minority of studies where there were significant differences in the ratings of men and women, the results suggested that women were rated higher than men.

Course-Related Factors

Research has consistently shown SETs are influenced by factors that are associated with the course content itself. These factors include: perceived difficulty of the course by students (Addison, Best, & Warrington, 2006) and whether the course itself is mandatory or an elective (Feldman, 1978; Gage, 1961; Lovel & Haner, 1955; Scherr & Scherr, 1990). In one study, researchers examined whether students were more apt to rate an instructor negatively when the course was more difficult than originally thought (Addison, Best, & Warrington, 2006). Regardless of the grade earned by participants in the study, the researchers predicted that students would rate their instructor more negatively if the course content was perceived as more difficult than they originally thought it would be. The researchers first found support for the fact that students who received higher grades evaluated their instructors more favourably than students who had received lower grades. Furthermore, it was found that, after controlling for the grade earned, students who thought the class was easier than expected than those who thought the class was harder rated instructors more favourably.

Some research has found that whether a class is a requirement versus an elective has no effect on SETs (e.g., Heilman & Armentrout, 1936; Hildebrand, Wilson, & Dienst, 1971), but a number of studies have found that this factor does actually have an effect (e.g., Feldman, 1978; Gage, 1961; Lovel & Haner, 1955; Pohlmann, 1975). Through analyzing student ratings of college teachers in the College of Education at the University of Illinois, Gage (1961) found evidence that instructors of elective courses were consistently rated more favourably than instructors of required courses. Similarly, in an in-depth analysis of college characteristics and college students' ratings of instructors, Feldman (1978) found some evidence suggesting that teachers of non-required or elective courses received higher ratings than required or mandatory

courses. The inconsistencies in findings with regards to the elective versus requirement debate are partly due to the fact that research in this area is dated and measurements used have varied across studies. A more recent analysis of student ratings based on whether a class is an elective versus a requirement is needed.

Student-Related Factors

A small amount of research has also been conducted on whether factors specific to students, such as student personality (Patrick, 2011), and personal interest in the subject matter (Marsh & Roche, 1997), have an effect on the outcome of SETs. For example, Patrick (2011) sought to determine whether any of the Big Five personality traits (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) had an effect on SETs. In her correlational analysis, it was found that one of the personality traits, agreeableness, was positively associated with the rating of an instructor's teaching effectiveness. In other words, these results indicate that as a student's level of self-reported agreeableness increases, so to does the instructor's level of teaching effectiveness.

The extent to which students find a particular course interesting or not has been found to have an impact on how favourably an instructor is rated (Marsh & Roche, 1997). In a comprehensive review of studies examining the relationships found between students' ratings and background characteristics, Marsh and Roche (1997) found that students who reported more interest in a class rated it higher than students who did not have a high level of interest. Specifically, the authors concluded that approximately one third of the expected grade effect was accounted for by prior interest in the subject matter. It is important to note that March and Roche were unable to determine if this level of interest existed before the start of the course or was generated by the course or the instructor.

Research in the area of possible extraneous factors that impact SETs is vast and still expanding. Some researchers have found that factors related to the instructor have an impact. These include the gender of the professor and their level of perceived physical attractiveness. Other researchers have found that factors associated with the course content itself plays a role in evaluations, including whether a course is a requirement or an elective course. Additionally, other researchers have found evidence of factors related specifically to the student also have an impact on SETs. These factors include personal interest in the course and the personality of the student.

The Importance of Grades

The relation between SET scores and student grades is one external factor that has received much attention in the literature (Kelly, 2012). In fact, what students expect to receive as grades or have already received as grades have been found to be one of the most consistent correlates of student ratings (e.g., Addison, Best, and Warrington, 2006; Greenwald & Gillmore, 1997; Krautmann & Sander, 1999; Olivares, 2001). For example, in a brief review of student evaluations on college teaching effectiveness, Wachtel (1998) found that there was a general consensus in the research literature that a moderate positive correlation existed between expected grade and student ratings. Students who expect higher grades tend to give more favourable ratings of their instructors. A number of possible explanations for this relation have been proposed by researchers (e.g., Greenwald & Gillmore, 1997; Marsh, 1987; Wachtel, 1998). For example, it could be that the presence of pre-existing differences among students has an impact on both teaching effectiveness and student evaluations (prior characteristic hypothesis). These could include differences in prior interest in the course subject, motivation levels, ability, and discipline (Arnold, 2009). Another explanation may be that their students rate effective

instructors more favourably because the high quality of teaching enables students to perform to their full potential (validity hypothesis). According to this view, a positive relation would support the validity of student evaluations (Greenwald & Gillmore, 1997). Another possible reason for the correlation between expected or received grades and SETs is the idea of grading leniency whereby instructors with more lenient grading standards receive more favourable ratings (grading leniency hypothesis; Greenwald & Gillmore, 1997).

Out of the three hypotheses described above, the grading leniency hypothesis has generated the most controversy in the field (Vaillancourt, 2013). According to this hypothesis, instructors have the ability to "buy" better evaluations by assigning higher grades (Marsh, 1987; Wachtel, 1998). Correlational research in this area has consistently found a low but positive correlation between expected and/or received grades and SETs, but causation cannot be inferred through correlation (Vaillancourt, 2013). Experimental research that has been conducted in this area has consistently shown strong support for the presence of the grading leniency hypothesis (e.g., Snyder & Clair, 1976; Vaillancourt, 2013; Worthington & Wong, 1979). Opponents of the grading leniency hypothesis have argued against the experimental research, stating that the research is weak and unrepresentative of what happens in naturally occurring situations (Marsh & Roche, 1997).

Another possible interpretation for the grades-ratings association can be found within *attribution theory* (Feldman, 1997; Gigliotti & Buchtel, 1990; Griffin, 2004). According to this theory, students may react in one of two ways if they receive a grade that differs from what was originally expected. If grades are lower than expected, students are likely to exhibit a defensive and self-protection mechanism commonly referred to as the self-serving bias (e.g., Arkin, Appelman, & Burger, 1980; Gigliotti & Buchtel, 1990; Heider, 1976; Snyder & Clair, 1976;

Zuckerman, 1979). In utilizing the self-serving bias, students attempt to protect their view of self and assign blame for the lower than expected grade to an external cause such as the instructor. If grades are higher than expected, students exhibiting the self-serving bias will tend to attribute this success internally and credit the successful grade to their intelligence, hard work, and ability rather than giving credit to the instructor's teaching effectiveness. Attribution theory implies that there is nonlinearity in predicted grade relations in that teachers may not be rewarded for giving high grades, but they may be punished for giving low grades.

Attribution research has generally established that there is an inherent self-serving bias in people's success/failure explanations (Ross & Fletcher, 1985). The presence of the self-serving bias has been found in a number of different events and settings, including employees receiving or being denied promotions, athletes performing well versus performing poorly, students achieving good or bad grades, and even automobile drivers narrowly missing an accident or getting into one (Shepperd, Malone, & Sweeny, 2008). The common thread among these different examples is that people view their positive or successful outcomes as primarily internally caused, but view their negative or failed outcomes as primarily externally caused. Shepperd and colleagues (2008) described internal causes as generally referring to "abilities, skills, personal traits, or effort, whereas external causes generally refer to the actions or inactions of others, luck, and circumstances such as the weather or economy" (pp. 895).

Research that has examined the prevalence of the self-serving bias has generally found the phenomenon to be relatively widespread and pervasive in the general population (Mezulis, Abramson, Hyde, & Hankin, 2004). Although significant differences exist to support the presence of the self-serving bias, it is unlikely that all individuals exhibit the bias in a similar fashion. Indeed, in their meta-analysis, Mezulis and colleagues (2004) concluded that the

prevalence of the self-serving bias demonstrated significant variability across factors such as age, culture, and psychopathology. Other researchers have found that individuals who suffer from social anxiety go so far as to reverse the self-serving bias (e.g., Arkin, Appelman, & Burger, 1980; Asendorpf, 1987; Teglasi & Hoffman, 1982), whereas those who display elevated levels of narcissistic personality traits exhibit the self-serving bias to the point that they are almost blind to any other point of view (e.g., Bushman & Baumeister, 1998; Vaillancourt, 2013). This great variability in the presence of the self-serving bias is cause for concern where SETs are involved due to the fact that they are used to inform high-status decisions, such as tenure and promotion (Gravestock & Gregor-Greenleaf, 2008; Kelly, 2012; Zabaleta, 2007).

The Self-Serving Bias and Students' Evaluations of Teaching

The presence of the self-serving bias in students receiving grades they either expected or did not expect to receive has been found across a number of studies (e.g., Arkin & Maruyama, 1979; Arnold, 2009; Davis & Stephan, 1980). In the late 1970s and early 1980s, a number of researchers examined the presence of the self-serving bias in undergraduate students when it came to exam performance. For example, Arkin and Maruyama (1979) sought to determine if successful students would assume greater personal responsibility for their performance in an undergraduate psychology class than would unsuccessful students. Consistent with the self-serving hypothesis, the authors found that successful students perceived internal factors as more important causes of their own performance and unsuccessful students perceived external factors as more important causes of their own performance. Similarly, research conducted by Davis and Stephan (1980) further supported the existence of the self-serving bias in SETs. In a sample of 127 undergraduate students from a large university in the United States, the authors found that the high expectations of students were based primarily on their perceptions that they possessed

the necessary amount of effort and ability to perform well. When their actual performances were consistent with these high expectations, participants emphasized ability and effort in explaining their success. Not surprisingly, those participants whose actual performance was quite low de-emphasized the importance of ability and effort and instead attributed their failure to the difficulty of the test. More recently, Arnold (2009) conducted research on whether examinations influenced students evaluations of instructors. Similar to past research, the author found that, among students who fail an examination, average evaluation scores were significantly lower after the exam. The author concluded that these results were compatible with a self-serving bias in student ratings, in which unsuccessful students externalize their failure but successful students do not externalize their success. Interestingly, the author concluded that the presence of the self-serving bias in SETs did not constitute a form of revenge against the instructors.

Past research has noted that the self-serving bias may serve important personal functions, including restoring damaged self-esteem or one's public self-image (Papps & O'Carroll, 1998). In fact, research dating as far back as the 1930s has found that individuals tend to seek a positive image of themselves and their environment with such vigour that reality is sometimes selectively interpreted or even patently ignored. According to Allport (1937, as cited in Mezulis, Abramson, Hyde, & Hankin, 2004) the self-serving bias may serve an adaptive function where it is used to protect egos from the harsh blows of reality. Some researchers have found evidence suggesting that the self-serving bias is a feature of the human cognition that is positively associated with physical and mental health. The presence of the self-serving bias has been associated with greater self-reported trait happiness (Rizley, 1978), less depression (Abramson & Alloy, 1981), better problem solving (Isen & Means, 1983), and better immune function (Taylor et al., 2000). An attenuated or absent self-serving bias has been associated with higher incidences of

depression (Sweeney, Anderson, & Bailey, 1986), poorer academic, work, and athletic performance (Peterson & Barrett, 1987; Seligman, Nolen-Hoeksema, Thornton, & Thornton, 1990), and poorer physical health (Lee & Seligman, 1997).

The Self-Serving Bias, Narcissism, and Student Evaluations of Teaching

Despite much research pointing to the fact that students who end up with lower than expected grades respond with lower ratings of their instructor; this form of self-serving bias is not often viewed as a form of revenge (e.g., Arnold, 2009). Research in the field of aggression has well established that social rejection, threats to ego, and negative feedback about one's abilities and characteristics may lead to anger and aggression (e.g., Barry, Chaplin, & Grafeman, 2006; Bushman & Baumeister, 1998; Bushman, Baumeister, Thomaes, Ryu, Begeer, & West, 2009; Jones & Paulhus, 2010; Konrath, Bushman, & Campbell, 2006; Stucke & Sporer, 2002; Twenge & Campbell, 2003). This area of research has examined the link between negative feedback and narcissism and some evidence has supported the fact that narcissism may be important in predicting who will aggress following negative feedback. Individuals who exhibit narcissistic traits are described as self-centered, manipulative, demanding, and having a grandiose sense of self-worth (American Psychiatric Association, 2013). The results from this area of research have consistently shown that individuals with inflated egos can, and often do, become aggressive when they are criticized or rejected by others.

In a study examining the links between narcissism, aggression, and self-esteem, Bushman and Baumeister (1998) sought to test the opposing predictions about the link between self-views and hostile aggression. The researchers conducted two studies to test the prediction that the combination of high narcissism and ego threat would lead to unusually high levels of aggression. In these two studies, participants were asked to write a one-paragraph essay on the controversial

topic of abortion and were led to believe that another participant would evaluate their essay. The essay was randomly assigned a positive or negative grade and participants were then able to aggress against their supposed evaluator through blasts of white noise in a competitive reaction time task. The authors found that individuals with high levels of narcissism were most aggressive following negative evaluations. More recently, Bushman and colleagues (2009) re-examined the data from their original study and found that the most aggressive individuals were those with the highest levels of narcissism. This finding provides further support for the threatened egotism hypothesis whereby the highest rates of aggression came from the combination of high scores on the trait of narcissism and ego threat.

Research conducted by Stucke and Sporer (2002) and Twenge and Campbell (2003) suggests that people with inflated and extremely positive self-views that are unstable and insecure might be more prone to aggression when their positive self-view is threatened by negative feedback. In an effort to expand on the experimental study conducted by Bushman and Baumeister (1998), Stucke and Sporer (2002) conducted two experimental studies to examine the relation between narcissism, anger, and aggression following ego-threat. The results of the study provided further evidence that people with an inflated, unstable self-view are more prone to angry and aggressive reactions after an ego-threatening event. In a series of four studies, Twenge and Campbell (2003) similarly hypothesized that people with higher levels of self-reported narcissistic personality traits would respond to social rejection with anger and aggression. The results of these four studies further complemented the research by Bushman and Baumeister (1998) and Stucke and Sporer (2002) in that individuals with elevated levels of narcissism were significantly more prone to respond to failure feedback with direct aggression.

The Self-Serving Bias, Shyness, and Student Evaluations of Teaching

The presence of the self-serving bias has been shown to be pervasive in the general population (Mezulis, Abramson, Hyde, & Hankin, 2004) and the research described above has generally found those who exhibit elevated levels of self-reported narcissism are more likely to exhibit the self-serving bias. Alternatively, some research has shown that shy or socially anxious individuals generally do not exhibit the self-serving bias at all (e.g., Alden, 1987; Arkin, Appelman, & Burger, 1980; Asendorpf, 1987; Franzoi, 1983; Shepperd, Arkin, & Slaughter, 1995; Teglasi & Hoffman, 1982). The term shyness is most often used to describe individuals who are reticent to engage in social interactions or who are socially withdrawn (Beidel & Turner, 1998). Shyness is the term most often used to label the emotional state of feeling anxious and inhibited in social situations (Cheek & Briggs, 1990). Alternative perspectives conceptualize shyness as a form of social anxiety (e.g., Buss, 1980; Crozier, 1995; Meleshko & Alden, 1993; Schlenker & Leary, 1982). Buss (1980) offered an ambitious theory on shyness, shame, embarrassment, and audience anxiety by stating that these can all be conceptualized as different forms of social anxiety, which he defined in terms of "discomfort in the presence of others...being upset or disturbed by others' scrutiny or remarks, or merely because others are present (p. 204). In addition, some researchers (e.g., Meleshko & Alden, 1993) describe shyness as a form of chronic social anxiety and suggest that these individuals are motivated by the desire to avoid negative social outcomes. For the purposes of this study, research that used the terms shyness and social anxiety were presumed to be referring to the same construct.

In the 1930s, Schouten (1935, as cited in Crozier, 1990) stated that "shyness is a phenomenon so universally human that we can easily say: someone who has never been shy or someone who, under certain circumstances, does not run the risk of becoming so, is an abnormal

person". Indeed, research does show that the prevalence rates of shyness are high, with one cross-cultural study reporting less than ten percent of respondents indicating they had never felt shy (Zimbardo, 1977). Some researchers have postulated that situational shyness as a transitory emotional state is a normal and functional aspect of human development and everyday adult life (Ford, 1987; Izard & Hyson, 1986). Some psychologists have argued that the positive connotations of shyness, such as modesty or sensitivity, should be emphasized (Aron, 1996; Gough & Thorne, 1986). Alternatively, there has been some research evidence that supports a negative image of shyness by documenting how shyness can be a barrier to personal well-being, social adjustment, and occupational fulfillment (e.g., Cheek & Briggs, 1990; Jones, Cheek, & Briggs, 1986).

Surprisingly, there has been little research examining the relationship between the self-serving bias and shyness. The research that does exist has consistently found that shy people typically tend to reverse the self-serving bias by accepting more personal responsibility for social failure than for success (e.g., Arkin, Appelman, & Burger, 1980; Teglasi & Hoffman, 1982). Individuals who describe themselves as shy tend to not experience an inflated sense of self-worth. Instead, they expect that their social behaviour will be inadequate and, as a result, they will be evaluated negatively by others (e.g., Capiocco, Glass, & Merluzzi, 1979; Leary, Kowalski, & Campbell, 1988; Smith & Sarason, 1975). Shyness has also been found to be associated with increased instances of remembering negative information about oneself (Breck & Smith, 1983; O'Banion & Arkowitz, 1977). Since shy people expect to be evaluated negatively, it would be assumed that they would welcome and appreciate positive feedback. Research has shown, however, that shy individuals more readily accept negative feedback over positive

feedback. In fact, shy individuals tend to resist and doubt the accuracy of the positive evaluations they receive (e.g., Alden, 1987; Asendorpf, 1987; Franzoi, 1983).

The Current Study

In light of the lack of agreement on the validity and usefulness of SETs, and in particular the extent to which the self-serving bias may impact these evaluations, the purpose of the study was two-fold. The first aim of the study was to explore the possible impact of student characteristics on the outcome of SETs utilizing an experimental approach. A significant majority of the past research examining SETs has been non-experimental and the few studies that have used experimental methods have consistently found participants to evaluate their supposed instructor more favourably in the high grade condition and more negatively in the low grade condition (e.g., Snyder & Clair, 1976; Vaillancourt, 2013; Worthington & Wong, 1979). The study adds to the SETs literature in the following ways. First, the study utilizes an experimental approach to examine the impact that the self-serving bias can have on SETs. This is an important consideration to make, especially given the fact that the results of SETs are often used to make important promotional and employment decisions. Second, a majority of the existing studies that examine the validity of SETs have not thoroughly considered the important role of student personality characteristics and it is only recently that researchers have begun to explore this in greater detail (e.g., Vaillancourt, 2013).

The second aim of the study was to consider the impact that the self-serving bias may have on the evaluations given to instructors. Following a content-analytic approach, consistent with *attribution theory* (Jones, et al., 1972; Weiner, 1974, 1986), the study sought to determine the extent to which students engaged in the self-serving bias in the context of written evaluations. Specifically, the study examined the self-serving bias in the specific contexts of self-reported

levels of narcissism and shyness and the extent to which these predicted whether students would exhibit characteristics of the self-serving bias. Previous research has consistently found that the self-serving bias is more prevalent among individuals who exhibit narcissistic traits (e.g., Bushman & Baumeister, 1998; Stucke & Sporer, 2002; Twenge & Campbell, 2003; Vaillancourt, 2013), but individuals who are shy tend to exhibit the opposite of the self-serving bias (e.g., Arkin, Appelman, & Burger, 1980; Teglasi & Hoffman, 1982). It is important to further examine the impacts that the self-serving bias can have on the validity of SETs, as the research in this area is minimal. Presently, however, the research does suggest that the self-serving bias can have a negative impact on the validity of SETs (e.g., Vaillancourt, 2013).

Research Questions and Hypotheses

Research question #1. Will the written feedback provided by students in the negative condition primarily consist of external reasons for their grade compared to primarily internal reasons for their grade if in the positive condition?

To date, there has been no research that has thoroughly explored the written evaluations of instructors by their students. The extent to which individuals internalize or externalize their grades can be further explored by reviewing written feedback. Previous research has found that individuals will be more apt to blame external sources for failure but attribute internal sources for success (e.g., Shepperd, Malone, & Sweeny, 2008). Accordingly, it was hypothesized that the written feedback students gave to their evaluators in the negative condition would primarily consist of external reasons for their low grade (*Hypothesis 1a*). Alternatively, it was predicted that the written feedback students gave to their evaluators in the positive condition would primarily consist of internal reasons for their high grade (*Hypothesis 1b*).

Research question #2. Will students who exhibit higher levels of self-reported narcissistic traits enact revenge through poor evaluations of their marker when in the negative condition and attribute their own skills and abilities when in the positive condition?

It has been well established in the research literature that most individuals will react with aggression when given poor feedback (Bushman & Baumesiter, 1998; Stucke & Sporer, 2002; Twenge & Campbell, 2003; Vaillancourt, 2013). This same research has suggested that individuals who exhibit higher levels of self-reported narcissistic personality traits will enact revenge when given a low grade through low SETs and will attribute a high grade to their own abilities. In the negative condition, it was expected that individuals with elevated levels of self-reported narcissism would be more likely to engage in the self-serving bias whereby they would externalize their low grade (*Hypothesis 2a*). Alternatively, research has proposed that individuals with elevated self-reported narcissistic traits will be more apt to attribute success internally when given praise, or in the case of the current study, given a high grade (e.g., Ross & Fletcher, 1985; Shepperd, Malone, & Sweeny, 2008). Consequently, in the positive condition, it was predicted that individuals with higher levels of self-reported narcissism would be more likely to engage in the self-serving bias whereby they would internalize their high grade (*Hypothesis 2b)*.

Research question #3. Will students who exhibit higher levels of self-reported shyness be more likely to internalize their failure if in the negative condition but refuse to credit their own skills and abilities if in the positive condition?

The little research that has examined the relationship between the self-serving bias and shyness has generally found that individuals who exhibit higher levels of shyness are less likely to attribute success to their own abilities and are more apt to internalize failure (e.g., Alden,

1987; Arkin, Appelman, & Burger, 1980; Asendorpf, 1987; Franzoi, 1983; Teglasi & Hoffman, 1982). Therefore, for those individuals in the negative condition, it was predicted that individuals with higher levels of self-reported shyness would be less likely to engage in the self-serving bias and would attribute failure internally (*Hypothesis 3a*). Alternatively, it was predicted that individuals in the positive condition would be more likely to attribute success externally (*Hypothesis 3b*).

Research question #4. Are students in the negative condition more likely to write longer, more detailed feedback to their evaluators compared to those in the positive condition?

Lastly, there is a common belief among the general public that individuals who are upset with a particular result may spend significantly more time expressing their displeasure. There is no question that individuals tend to be more verbose in their defense. If you look at the reviews for articles, books, art, movies, and so on, you will be hard pressed to find more positive reviews than negative. In line with this view of common thought is research that has shown that "bad is stronger than good" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). The authors found that bad impressions and negative stereotypes are quicker to form and more resistant to disconfirmation than good ones. As a result, it was predicted that individuals in the negative condition would write longer, more detailed comments about their evaluator than individuals in the positive condition (*Hypothesis 4*).

Methods

Participants

Participants included 563 (299 women, 263 men, 1 unreported) predominantly first-year (84.8%) university students (M_{age} = 18.84 SD = 2.55) from a large south-western Ontario university. Students were recruited from the psychology participant pool and given course credit

for participation. Students came from a number of different faculties, with the most prevalent being Science (42.6%), Social Science (26.6%), and Business (13.2%). The cultural/ethnic distribution of the sample was predominantly Caucasian (46.8%), followed by South Asian (18.6%), and Asian (17.5%). The majority of participants were not majoring in Psychology and did not intend to major in Psychology in the future (78.9%). Participants were excluded if English was not the language they primarily communicated in.

Measures

Narcissism. Self-reported levels of narcissism were reported using the *Narcissistic Personality Inventory* (NPI; Raskin & Terry, 1988). The NPI is a forty-item forced-choice questionnaire used to measure symptoms of narcissism in the general population. Participants are asked to choose between competing items, such as "People sometimes believe what I tell them" versus "I am an extraordinary person". Scores can range between zero and 40, with average scores between 16 and 19 (Raskin & Terry, 1988). The items in this questionnaire are based on the diagnostic criteria for Narcissistic Personality Disorder in the DSM-IV-TR (APA, 2000), although those who score high on this measure do not necessarily meet the criteria for a clinical diagnosis of narcissism. The NPI is one of the most common measures used by researchers to assess levels of narcissism in the general population. This measure has consistently been found to have high internal consistency ($\alpha = .78$) and has been tested against various measures of narcissism (Raskin & Hall, 1979, 1981; Raskin & Terry, 1988; Watson, Grisham, Trotter, & Biderman, 1984).

Shyness. Self-reported levels of shyness were reported using the *Shyness and Sociability Questionnaire* by Cheek and Buss (1981). The Shyness and Sociability Questionnaire is a 13-item questionnaire in which respondents are asked to determine the extent to which each item

describes their character. The questions are answered on a 5-point Likert-type scale ranging from 0 (not at all characteristic) to 4 (extremely characteristic). Sample items include: "I feel tense when I'm with people I don't know well" and "I am often uncomfortable at parties and social functions". The Shyness and Sociability Questionnaire has been found to have high internal consistency ($\alpha = .85$) and is often used to measure shyness and sociability in the general population (Arkin & Grove, 1990; Asendorpf & Meier, 1993; Bruch, Gorsky, Collins, & Berger, 1989; Cheek & Buss, 1981; Czeschlik & Nurk, 1995).

Student Evaluations of Teaching. Participants were given a questionnaire developed by Vaillancourt (2013) where they were asked four questions about their evaluator, three of which were on a 7-point Likert-type scale ($\alpha = .86$). The first question asked participants to rate the overall quality of the evaluator's marking ability (1 = "very poor", 7 = "excellent"). The second question asked participants to determine the extent to which they felt the evaluation procedures were fair (1 = "disagree very strongly", 7 = "agree very strongly"). The third question asked participants to determine if they felt the evaluator's feedback was helpful (1 = "disagree very strongly ", 7 = "agree very strongly"). Finally, participants were provided with an opportunity to provide a written comment about their evaluator's marking abilities, which were evaluated in this study.

Procedure

Participants were informed that they would be participating in two related studies. The purpose of the first study was to investigate the evaluation skills of either professors or teaching assistants and the purpose of the second study was to examine whether "people with different personalities" held differing opinions about euthanasia. This cover story was used to help ensure that participants did not become suspicious about the true nature of the study and to avoid unduly

influencing SETs ratings. The true purpose of the study was to determine if their self-reported levels of narcissism and shyness predicted the extent to which students blamed their instructor if given a low grade on their assignment.

Students were informed of their rights as participants, including the right to confidentiality and the right to withdraw from the experiment at any time. Once consent was obtained from the participants, they were asked to type out a short paragraph essay (between 150-250 words) defending their point of view on the topic of euthanasia. A standard definition of euthanasia was provided and participants were informed that they would have 20 minutes to type the essay on a laptop computer. Upon completion of the essay (or if the 20 minute time-limit was up) participants were informed that their essay would be marked by a student teaching assistant or professor (random). They were informed that their essay would be marked on organization, originality, writing style, clarity of expression, persuasiveness of arguments, and overall quality. In actuality, the participants were assigned a standard grade that was not dependent on the quality of their essay. Half of the participants were given a poor evaluation whereas the other half was given a positive evaluation.

Participants were randomized into two conditions. In the negative feedback condition (n = 128 men; n = 144 women; 1 unreported), they were given standardized numeric feedback on their essay that ranged from C- on originality and persuasiveness of arguments to a D on all other evaluation criteria. They were also given the written comment, “This is one of the worst essays I have ever read!” Participants in the positive condition (n = 135 men; n = 155 women), were given standardized numeric feedback on their essay that ranged from B+ on organization to A+ on originality and were given the written comment “No suggestions, great essay!” In both conditions, comments were added to each essay at approximately the same location to give the

impression that the essay was actually read by the evaluator (it was not). During the time that their essay was being "marked", participants were asked to complete a series of questionnaires examining levels of narcissism, shyness, and demographic information. Participants were then provided with the feedback on their essay. Following this, they were asked to fill out an evaluation form pertaining to the evaluator's marking abilities, including a section allowing for written feedback. Participants were assured that their anonymity would be strictly guarded and that the evaluators would have no way of knowing who wrote the essay. In addition, the evaluations provided by students regarding their evaluators would be anonymous.

At the conclusion of the study, participants were given a debriefing form and were verbally told that their essay had not really been evaluated and that they were actually randomly assigned to a standard positive or negative evaluation. The debriefing sheet revealed to them that the real purpose the study was to examine personality styles in relation to positive or negative feedback. In addition, participants were provided with reasons regarding why the concealed methodology was necessary for the study.

Coding Procedures

Borrowing from research on the self-serving bias and attribution theory, a content analysis was conducted on the written evaluations given to instructors. First, responses to the item, "how would you rate your evaluator…" were transcribed verbatim and checked for accuracy. Two independent raters coded the responses of participants with regards to whether they fit into one of three attribution categories: (1) *external* (e.g., "…personally, I cannot say it is the easiest to write an essay within 20 minutes without concrete facts"), (2) *internal* (e.g., "I think I'm a fairly decent writer, being in Humanities and all…"), or (3) *neutral* (e.g., "No suggestions, great evaluation!"). The determination of whether a response was internal, external,

or neutral was conceptually driven and based on previous research in attribution theory and the self-serving bias (e.g., Arkin, Appelman, & Burger, 1980; Gigliotti & Buchtel, 1990; Heider, 1976; Snyder & Clair, 1976; Zuckerman, 1979). Coding disagreements were settled with a "blind" third independent rater. Cohen's kappa for inter-rater agreement was found to be .87.

Within each of these three attribution categories, a total of 12 reasons for an essay grade were developed. Some of these coding categories were conceptually driven and developed a priori on the basis of previous research in attribution theory and the self-serving bias (e.g., Arkin, Appelman, & Burger, 1980; Gigliotti & Buchtel, 1990; Heider, 1976; Snyder & Clair, 1976; Zuckerman, 1979). The external categories that were conceptually driven and developed a priori were: (1) *issue of time* (e.g., "20 minutes is not long enough to write an essay about a subject that is very controversial"), (2) *questioning competence of evaluator* (e.g., "I think the evaluator failed in his/her competence when marking this paper"), (3) *lack of resources or prior knowledge on subject* (e.g., "…participants did not have any previous knowledge, other than what they've heard, on the topic"), and (4) *unfair marking on essay* (e.g., "I believe the evaluator marked my essay very unfairly"). The internal category that was conceptually driven and developed a priori was: (1) *participant not a good writer* (e.g., "I have always been bad at writing…"). Lastly, the neutral categories that were conceptually driven and develop a priori were: *and* (2) *helpful feedback provided* (e.g., "…not only graded my paper but he/she provided comments…"). The other coding categories were developed via emergent coding following a preliminary examination of the data. The external categories that were developed via emergent coding were: (1) *bias in marking* (e.g., "…the evaluator had a different take on the subject being discussed, and he/she took a biased approach to marking…"), (2) *revenge toward evaluator* (e.g., "if I'm the worst essay writer ever, than this evaluator was the worst evaluator ever"), and (3) *rushed*

marking (e.g., "…seems a bit rushed in marking"). The internal categories were developed via emergent coding were: (1) *agree with competence of evaluator* (e.g., "the evaluator is competent in terms that he/she knew about the need for clarification of many points…"), and (2) *confident in writing abilities* (e.g., "I feel the quality of my writing is superb…"). Coding disagreements were settled with a "blind" third independent rater. Cohen's kappa for inter-rater agreement was found to be .82.

Results

To determine the percentage within each coding category, a frequency table was constructed, with participants being grouped based on whether they were in the negative or positive condition. Following this, a chi-square analysis was conducted to determine if the proportion of attribution (external, internal, or neutral) varied based on whether participants were grouped into the negative or positive condition. To determine the extent to which participants engaged in the self-serving bias and whether the personality variables of narcissism or shyness had an impact, a multinomial logistic regression analysis was conducted. Lastly, an independent groups t-test was conducted to determine if participants in the negative condition used more words in their written feedback compared to those in the positive condition.

Attribution Categories

The observed frequencies for attribution are presented in Table 1. Out of the total sample size (n = 563), 35.3% externalized their grade, 5% internalized their grade, and 59.7% were neutral toward their essay grade. For those participants randomized into the negative condition (n = 273), 45.1% externalized their essay grade, 5.5% internalized their grade, and the remaining 49.5% were neutral toward their grade. For those participants randomized into the positive

condition (n = 290), 26.2% externalized their essay grade, 4.5% internalized their grade, and 69.3% were neutral toward their essay grade.

Low Grade (Negative Condition)

The observed frequencies for each of the coding categories in the negative condition are presented in Table 1. Students were able to provide multiple reasons in their written feedback, with the total number of reasons ranging from one to five. Close to half of the participants felt that they lacked sufficient time (40.7%) to complete the essay (e.g., "…I was given 20 minutes to do it, so I wasn't able to write up something worthwhile"). A small percentage of students (3.7%) felt that their evaluator was biased in marking their essay (e.g., "…the evaluator had a different take on the subject being discussed and he/she took a biased approach to marking") and some (1.8%) went so far as to enact revenge on their evaluator through their written comments (e.g., "…this evaluator is a complete dumb ass, excuse my french"). Some students (8.4%) questioned whether their evaluator was competent enough to mark their essay (e.g., "…it is evident that my evaluator does not read many essays and is therefore not qualified to be marking mine"), whereas some (9.9%) felt that the time it took to grade their essay was rushed (e.g., "it seems like the evaluator rushed the evaluation"). A very small percentage (3.3%) of participants in the negative condition felt that they were a bad writer, which contributed to their low grade (e.g., "I have always been bad at writing"). Approximately one quarter of participants (24.5%) did find that their evaluator was competent in their marking (e.g., "the essay was unstructured and poorly worded, so I agree with the marking"). The majority (77.3%) felt that the feedback provided was vague and lacked detail (e.g., "the feedback on the essay itself was very vague and the amount of feedback didn't really match the grades the marker gave"). Some participants (13.2%) felt that they lacked sufficient resources or prior knowledge on the subject matter to

complete the essay properly (e.g., "…participants did not have any previous knowledge other than what they've heard"). Eleven percent of participants argued that the entire marking procedure was unfair (e.g., "…the marker did not elaborate on any of their critiques and was very unfair with their final comments"). A very small proportion of participants (1.1%) felt confident in their writing abilities (e.g., "I have confidence in what I write and have never been told one of my essays is a failure"). Lastly, a small percentage of participants (2.9%) found the evaluator's feedback to be helpful (e.g., "using feedback such as elaborate and needs clarification are useful").

High Grade (Positive Condition)

The observed frequencies for each of the coding categories in the positive condition are presented in Table 2. Similar to the negative condition, students were able to provide multiple reasons in their written feedback, with the total number of reasons ranging from one to five. Less than a quarter of the participants (22.8%) felt that the time constraint was a factor in their essay and some went so far as to state the evaluator took the issue of time into consideration with their marking ("e.g., I agree with the evaluator's marking of my essay given the time restrictions placed upon while writing it"). A very small proportion of participants (2.1%) did feel their evaluator was biased (e.g., "I felt that the marker was positively biased towards my opinion, and this positively affected the grade is received"). None of the participants enacted revenge on their evaluator through written comments. Some participants (21.4%) did question the competence of their evaluator (e.g., "the evaluator is not a very good marker, I received a mark of an A when I didn't even complete the essay") and a small percentage (7.6%) felt that the marking was rushed (e.g., "the fact that there were no suggestions made it seem as if the professor was in a huge rush…"). A small proportion (2.4%) felt that they were a bad writer (e.g., "I like to be

evaluated…because I'm not good at writing"). Almost half of the participants (43.4%) felt that their evaluator was competent in their marking abilities (e.g., "…the evaluator's marking abilities are great in my opinion and I feel the marks I received were fair and deserved"). Similarly, more than half of the participants (60%) felt that the comments left by the evaluator were vague and lacked detail (e.g., "…it would be very helpful if there were any suggestions about organization"). A small percentage (2.8%) did not feel they had the appropriate resources or background knowledge to complete a sufficiently good essay (e.g., "this essay was pretty garbage because no research was completed, and I have never thought about this topic"). A small percentage (1.7%) indicated that the evaluator was vague or lacked detail in their marking (e.g., "there was no real feedback, other than good comments, but nothing to help improve the essay"). Similarly, a small percentage (1.7%) felt confident in their writing abilities (e.g., "…I feel the quality of my writing is superb…"). Lastly, some participants (8.6%) found the feedback provided by the evaluator to be helpful (e.g., "the marking abilities of the evaluator are very good, they gave appropriate feedback when needed…").

Externalizing versus Internalizing

A chi-square test was performed to determine if the proportion of attribution (internal, external, or neutral) varied by condition (positive or negative). A summary of the chi-square frequencies is presented in Table 3. The extent to which participants internalized or externalized their essay grade was significantly related to the condition they were randomized into, χ^2 (2) = 23.72, p = .000. For those participants that externalized their essay grade, 21.8% were in the negative condition and 13.5% were in the positive condition. For those participants that internalized their essay grade, 2.7% were in the negative condition and 2.3% were in the positive

condition. Finally, 24% of participants in the negative condition were neutral toward their essay grade versus 35.7% in the positive condition.

Self-Serving Bias as Predicted by Self-Reported Levels of Narcissism and Shyness

A multinomial logistic regression analysis was conducted to predict the extent participants engaged in the self-serving bias (i.e., 3 outcomes – neutral, internalize, externalize) using condition (positive/negative), self-reported levels of narcissism and shyness, as well as interactions between condition and both narcissism and shyness were entered as predictors. The reference category was neutral whereby participants did not internalize or externalize their essay grade. The effects of shyness and narcissism did not vary by condition, so the interactions were dropped from further analysis. The extent to which individuals reported levels of shyness predicted whether participants would internalize their essay grade was not significant, χ^2 (2) = 5.00, p = .08. Shyness did distinguish between internal and neutral, b = 0.66, Wald χ^2 (1) = 4.56, p < .05. Self-reported levels of narcissism did not significantly predict the extent to which participants internalized their essay grade, b = 1.26, Wald χ^2 (1) = 0.81, p = .40.

Word Count

There was a statistically significant difference in the word count for those in the positive condition (M = 50.15, SD = 30.92) and the negative condition (M = 60.15, SD = 40.96); t(505.42) = 3.25, p < .001 whereby those in the negative condition were more verbose than those in the positive condition.

Discussion

One of the purposes of the study was to examine whether the extent to which students externalized or internalized a high or low essay grade predicted whether they engaged in the self-serving bias. In their meta-analysis, Mezulis and colleagues (2004) determined that the

prevalence of the self-serving bias in North America was high, indicating that individuals would be more likely to engage in the self-serving bias in a wide variety of situations. To explore this further, we conducted a content analysis on the written comments a group of students wrote regarding their evaluator. We predicted that the students in the negative condition would write primarily external reasons for their low grade compared to those in the positive condition who would write primarily internal reasons for their high grade.

Additionally, we were interested in examining whether students' self-reported levels of narcissism had an impact on whether students would be more apt to engage in the self-serving bias by externalizing their low grade or internalizing their high grade. Alternatively, we wanted to determine whether students' self-reported levels of shyness had an impact on whether students would be less likely to engage in the self-serving bias by internalizing their low grade and externalizing their high grade. Lastly, we aimed to further explore the adage that people would be more verbose when giving negative feedback..

Explanations for Essay Grades

Since our study was one of the first to explore the written feedback given by students to their evaluators, we sought to explore the different reasons students used to explain their high or low essay grade. Consistent with attribution theory and the self-serving bias (e.g., Arkin & Maruyama, 1979; Arnold, 2009; Davis & Stephan, 1980; Shepperd, Malone, & Sweeny, 2008), it was hypothesized that a majority of the reasons for a high essay grade would be internalized and students would attribute success to their own skills and abilities. Alternatively, it was hypothesized that a majority of the reasons for a low essay grade would be externalized and students would be more apt to place blame elsewhere. A content analysis was conducted on the

written feedback to further examine the reasons students attributed for the high or low grade they received on their essay.

Our results indicated that a very low number of students internalized their essay grade in the positive condition, even less than those in the negative condition. This was an unexpected finding given our predictions. One possible explanation for this finding could be linked to the emphasis that students generally tend to place on grades (e.g., Crocker, Kapinski, Quinn, & Chase, 2003). Crocker and colleagues (2003) have pointed out that grades for course work are one of the most important events for university students. The consistently positive correlation between expected grades and SETs further supports this fact (e.g., Addison, Best, and Warrington, 2006; Greenwald & Gillmore, 1997; Krautmann & Sander, 1999; Olivares, 2001). The fact that so few participants internalized their grade also calls attention to the issue of grading leniency (e.g., Greenwald & Gillmore, 1997; Marsh, 1987; Wachtel, 1998). If participants are placing such emphasis on external reasons for their low grades, it is possible to speculate that they would be less likely to do this if instructors gave them high grades. The grading leniency hypothesis proposes that instructors have the ability to "buy" better evaluations by giving students higher grades, even if such grades are not warranted (Wachtel, 1998).

Students who were randomized into the high essay grade condition voiced a wide array of opinions as a way to explain their high grade. A number of students felt that their evaluator was competent in their marking abilities but that the comments left were vague and lacked detail. Research examining students' perceptions of feedback has found the congruence of feedback to be one of the most important factors to students (Poulos, Mahony, 2008). If the feedback given is not in line with the expectations of students, then it is likely that the credibility of such feedback will be questioned. Some students who either externalized or internalized their grade

did feel confident in their writing abilities, but this was a small percentage overall. Research in the area of self-protection (e.g., Martin, Marsh, & Debus, 2001) has found that students to protect their self-worth often use strategies such as self-handicapping and defensive pessimism. Specifically, these researchers found that students were more apt to proactively dissociate themselves from potential failure whereby they externally attribute their performance. Given the parameters of the task of writing an essay in 20 minutes with no outside resources, it is reasonable to predict that many felt they did not do well on the task and had reasoned out explanations for this ahead of receiving their feedback.

The predetermined, small window of time was cited as the most common reason students who were randomized into the negative condition felt they did not do well on the essay. A number of the students who externalized their grade also stated that they did not have enough background knowledge or access to external resources to write a "good quality" essay. Additionally, students who externalized their grade were more apt to question the evaluator's level of competence in grading their essay, effectively placing blame on the evaluator rather than themselves. .Previous research has found that individuals generally tend to respond with aggression when given poor feedback (Bushman & Baumeister, 1998; Stucke & Sporer, 2002; Twenge & Campbell, 2003; Vaillancourt, 2013). Previous research has also shown a consistently positive correlation between expected grades and evaluations (Wachtel, 1998) and a majority of students in this study expected the evaluator to be more lenient in their marking given the parameters of the assignment (e.g., short amount of time and lack of outside resources). When this was not the case for students in the negative condition, many responded by evaluating their marker negatively with regards to their level of competence.

Predicting the Presence of the Self-Serving Bias

Researchers have previously found evidence supporting the fact that individuals with elevated levels of narcissism would be more likely to attribute success internally and more likely to attribute failure externally (e.g., Barry, Chaplin, & Grafeman, 2006; Bushman & Baumeister, 1998; Bushman, Baumeister, Thomaes, Ryu, Begeer, & West, 2009; Jones & Paulhus, 2010; Konrath, Bushman, & Campbell, 2006; Stucke & Sporer, 2002; Twenge & Campbell, 2003). Following from the viewpoint that the self-serving bias is present in a large variety of settings and used to protect one's ego, several hypotheses were put forth. We expected that students with elevated levels of self-reported narcissistic traits would be more likely to take credit for their high essay grade and consequently internalize their success. We also predicted that students with higher levels of self-reported narcissistic traits would be more apt to engage in the self-serving bias and consequently blame the instructor for the low grade on their essay. These hypotheses were not supported in that no differences were found in self-reported levels of narcissism for students who externalized or internalized their essay grade.

One explanation for finding no differences in self-reported levels of narcissism could be that such a small proportion of the overall sample that internalized their essay grade. If the differences in levels of self-reported narcissism were not large enough to detect with such a small percentage, then no differences between individuals that internalized versus not internalized would be found. Another possible interpretation stems from the observation that almost a quarter of students (24.5%) in the negative condition and close to half (43.4%) in the positive condition felt that their evaluator was competent in their marking abilities. A majority of the sample in the study were first year students (84.8%) and this is a time of transition and adjustment for many students (e.g., Chemers, Hu, & Garcia, 2001; Hillman, 2005). Students

who are in their first year may be more likely to defer to authority more often than students in upper year levels (e.g., "since he or she is a professor, I assume he must be well informed as to what a deserving essay is…"). Even if a first year student does exhibit higher levels of self-reported narcissism, it could be that the transition to university and deference for authority at this level may overshadow the need for blame.

Previous researchers have found that individuals who display elevated levels of shyness are less likely to attribute success internally and would be more likely to attribute failure internally (e.g., Alden, 1987; Arkin, Appelman, & Burger, 1980; Asendorpf, 1987; Franzoi, 1983; Shepperd, Arkin, & Slaughter, 1995; Teglasi & Hoffman, 1982). We predicted that students with higher levels of self-reported shyness would be less likely to engage in the self-serving bias and internalize their low grade or externalize their high grade. These hypotheses were partially supported. While no differences were found between students' levels of self-reported shyness among those who externalized their essay grade, levels of self-reported shyness did distinguish between students who internalized their grade versus those who were neutral toward their grade. Students who reported higher levels of shyness were more likely to internalize rather than feel neutral toward their essay grade.

Previous research on shyness has generally found shyness to be universal and stable over time (Gest, 1997). Individuals who exhibit higher levels of self-reported shyness are generally shy across different situations and the incidence of shyness in the general population has been found to be quite high (Zimbardo, 1977). Despite the small percentage of participants who internalized their essay grade compared to those who were neutral, the pervasiveness of this personality trait is likely why a difference was found. Individuals who describe themselves as shy are more likely to expect and internalize failure (e.g., Henderson, 2002; Shepperd, 1995).

The negative feedback for those students that were randomized into the negative condition was likely harsh enough for them to internalize their grade and blame themselves.

The most common reason cited by students who externalized their grade was the lack of time they felt they had to complete the essay (40.7% in the negative condition and 22.8% in the positive condition). The 20-minute time limit on the essay is a valid concern and would likely be cited as an explanation despite self-reported levels of shyness. In addition, some participants reported that they lacked the prior knowledge or resources to sufficiently complete the essay, which is a valid statement given that participants were not provided with access to outside resources to complete the essay. The truth of this reason may mean that students would be more likely to state this in their feedback despite their self-reported levels of shyness.

Extent of Written Feedback

To date, this is the first study to explore the relationship between the type of evaluation given (positive versus negative) and the word count of written feedback. As predicted, there was a significant difference in the number of words used in written feedback whereby students in the negative condition used significantly more words than those in the positive condition. This suggests that individuals will spend more time giving feedback to their evaluators when given a low grade and is consistent with the common belief that people are more verbose when giving negative feedback versus positive feedback. This finding provides an important first look into the relevance of written feedback and further supports the existence of the self-serving bias in SETs. If students are given a high grade, it could be that they are likely attributing this success internally and thus feel no need to spend much time on evaluating their instructor. Alternatively, if they receive a low grade, they could be attributing this low grade externally, and who best to blame than the evaluator who was “responsible” for giving them this low grade?

Implications

The current study has implications for the future use and interpretation of SETs among postsecondary institutions. As the consistently growing research has clearly shown, the use of SETs as a primary means to evaluate teaching effectiveness will not be going anywhere anytime soon. In addition, the emphasis that many postsecondary institutions place on SETs with regards to tenure and promotion of instructors necessitates further review surrounding the validity of SETs (Gravestock & Gregor-Greenleaf, 2008; Kelly, 2012; Zabaleta, 2007). Faculty continue to express concern over the importance administrative staff place on these measures (McKeachie, 1997; Zabaleta, 2007) and the trend of grade inflation and "buying" favourable SETs with high grades will only continue to expand (Greenwald & Gillmore, 2007).

The results of this study seem to suggest that the balance between grades and feedback is one that is complex and multifaceted. In the present study, students were "marked" on several different areas and written feedback was provided in multiple areas throughout the essay. Recent research in this area has found the feedback was generally ignored and the focus was primarily on the final grade (e.g., Vaillancourt, 2013). When the written feedback was examined in further detail in this study, it was found that many students found the feedback provided by the evaluators was vague, lacked detail, and was generally unhelpful. This would seemingly contradict the evidence suggesting that grades matter more than feedback and gives an introductory glimpse into the complexities that instructors face when grading the work of their students.

To date, research has focused primarily on course and instructor related factors with little consideration of the role that student characteristics can play in SETs. Given the importance that students place on grades, it is important to further examine the role that student characteristics

can play in the context of final grades and feedback. This study was one of the first to examine the role that student personality plays in the written feedback of SETs and provides a glimpse into the complexities that exist. Despite finding no differences in the presence of self-serving biases within the context of self-reported levels of narcissism and shyness, further research is warranted. Our results did find that students were more likely to externalize their essay grade if they were randomized into the negative condition. If students are engaging in self-serving biases when filling out SETs, then the extent to which the written feedback can be interpreted must be questioned. Students may be more likely to write extensive feedback if they are given a low grade versus a high grade, resulting in a potentially unfair assessment of the true capabilities of the instructor.

Strengths, Limitations, and Future Directions

Strengths

To our knowledge, this is the first study to utilize a content analytic approach to examine the presence of the self-serving bias in SETs. Taking into account the role that personality characteristics can play in the evaluations that students give instructors is an important area of research and this study is an important first step toward examining this in further detail. The in-depth analysis of the feedback provided by the students allows us to examine the presence of the self-serving bias in further detail. The experimental design and randomized conditions used in this study ensure that causation can be inferred and valid conclusions can be drawn. In addition, the design of the study ensures that future researchers will be able to replicate the methodology. The use of SETs as a measure of aggression rather than white noise as has been used in past research (e.g., Bushman & Baumeister, 1998) is more valid given that most adults aggress against others using indirect aggression (Vaillancourt, 2005). Utilizing a content analytic

approach to examine the written feedback of the participants provides us with further insight into the reasons commonly used by students to explain their grades.

Limitations

Vaillancourt (2013) recently stated that those who oppose the grading leniency hypothesis would argue the results from the study do not generalize to the real world. Students have an expectation that the grade they receive will reflect past and current efforts. In the context of the current study, this is not the case. At the same time, however, the fact that participants received grades that went against their expectations could be in line with the inflated expectations often seen among students entering their first year of postsecondary studies (Kruger & Dunning, 1999). This area of research has found that it is very common for first-year students to enter college or university with inflated expectations of how well they will do. All too often students are shocked when their final grades do not reflect their perceived ability and knowledge.

The randomized condition of this study necessitates that some well-written essays were unfairly marked. For example, it could be that the quality of the essay written by some participants was strong and their response to receiving a low grade was legitimate. Alternatively, it could be that the quality of the essay written by some participants was quite low but they still externalized their grade, providing further support for the presence of the self-serving bias. An important next step in this area of research would be to objectively mark each of the essays in an effort to further examine the pervasiveness of the self-serving bias.

SETs are typically given before students receive their final grade. As we move toward an ever-expanding online presence, it is hardly surprising that many institutions now offer SETs online and these forms can be accessed before and after the final exam (Avery, Bryant, Mathios, Kang, & Bell, 2006). Some research has shown that when SETs are completed does matter

(Arnold, 2009). In this study, the researcher found no observable differences in SETs filled out before or after the final exam among those who passed. There were, however, significant differences in the SETs among students who failed the exam in that those who failed and provided feedback after the final exam consistently submitted lower evaluations.

Future Directions

The use of SETs in postsecondary institutions needs to be re-evaluated and factors related to student characteristics and personality should be taken into consideration when evaluating teaching performance. Research in social psychology has long considered the importance of person in context interactions (e.g., Magnusson & Stattin, 2006; Neufeld, Rasmussen, Lopez, Ryder, & Magyar-Moe, 2006; Vaillancourt, 2013). It is interesting to note that research to date in the area of SETs has not given much thought to person in context interactions and future research is needed.

Recent research (e.g., Vaillancourt, 2013) has demonstrated that the feedback given to students does not seem to matter as much as the final grade. This research found that, despite buffering the effects of a low grade with praise, students would still give low ratings to their evaluators. These findings provide further support for the grading leniency hypothesis. Given that a majority of the feedback for evaluators in this study emphasized the need for more detailed feedback, it would be pertinent to examine the importance of feedback further, especially in the context of student personality characteristics.

Conclusion

When integrating the findings from the present study, we found some support that self-reported levels of shyness did play a role in whether participants internalized their grade, but more research (both quantitative and qualitative) is needed to further understand the complex

relationship between student personality characteristics and SETs. There is a need to re-examine the use of SETs as the only means to evaluate the teaching performance of instructors. Given that SETs are, and will continue to be, an important measure used to evaluate teaching effectiveness, the role that student personality characteristics can play in these SETs is an important consideration to make. To date, the focus of SETs research has primarily focused on factors specific to the instructor and the course without considering the complex role of student characteristics. Consequently, the use of SETs in the context of administrative decisions must consider the contribution that the characteristics of students make. In addition, given that students are much more likely to externalize a poor grade, it is important to consider the propensity of instructors to be lenient in their grading in an effort to encourage higher evaluations. The role of grading leniency in SETs is an important one to consider and requires further attention moving forward.

References

Abramson, L. Y., & Alloy, L. B. (1981). Depression, nondepression, and cognitive illusions: Reply to Schwartz. *Journal of Experimental Psychology, 110*(3), 436-447. doi:10.1037/0096-3445.110.3.436

Addison, W. E., Best, J., & Warrington, J. D. (2006). Students' perceptions of course difficulty and the ratings of the instructor. *College Student Journal, 40*(2), 409-416.

Alden, L. (1987). Attributional responses of anxious individuals to different patterns of social feedback: Nothing succeeds like improvement. *Journal of Personality and Social Psychology, 52*(1), 100-106. doi:10.1037/0022-3514.52.1.100

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev). doi: 10.1176/appi.books.9780890423349

Arkin, R. M., Appelman, A. J., & Burger, J. M. (1980). Social anxiety, self-presentation and the self-serving bias in causal attribution. *Journal of Personality and Social Psychology, 52*(1), 100-106. doi:10.1037/0022-3514.52.1.100

Arkin, R. M., & Grove, T. (1990). Shyness, sociability, and patterns of everyday affiliation. *Journal of Social and Personal Relationships, 7*(2), 273-281. doi:10.1177/0265407590072008

Arkin, R. M., & Maruyama, G. M. (1979). Attribution, affect, and college exam performance. *Journal of Educational Psychology, 71*(1), 85-93. doi:10.1037/0022-0663.71.1.85

Arnold, I. J. (2009). Do examinations influence student evaluations? *International Journal of Educational Research, 48*(4), 215-224. doi:10.1016/j.ijer.2009.10.001

Aron, E. N. (1996). *The highly sensitive person.* New York, NY: Birch Lane Press.

Asendorpf, J. B. (1987). Videotape reconstruction of emotions and cognitions related to shyness. *Journal of Personality and Social Psychology, 53*(3), 542-549. doi:10.1037/0022-3514.53.3.542

Asendropf, J. B., & Meier, G. H. (1993). Personality effects on children's speech in everyday life: Sociability-mediated exposure and shyness-mediated reactivity to social situations. *Journal of Personality and Social Psychology, 64*(6), 1072-1083. doi:10.1037/0022-3514.64.6.1072

Association of Universities and Colleges of Canada (2012). Retrieved from http://www.aucc.ca

Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *The Journal of Economic Education, 37*(1), pp. 21-37. doi:10.3200/JECE.37.1.21-37

Barry, C. T., Chaplin, W. F., & Gafeman, S. J. (2006). Aggression following performance feedback: The influence of narcissism, feedback valence, and comparative standard. *Personality and Individual Differences, 41*(1), 177-187. doi:10.1016/j.paid.2006.01.008

Beidel, D. C., & Turner, S. M. (1998). *Shy children, phobic adults: Naure and treatment of social phobia.* Washington, DC: American Psychological Association.

Breck, B. E., & Smith, S. H. (1983). Selective recall of self-descriptive traits by socially anxious and nonanxious females. *Social Behavior and Personality, 11*(2), 71-76. doi:10.2224/sbp.1983.11.2.71

Bruch, M. A., Gorsky, J. M., Collins, T. M., & Berger, P. A. (1989). Shyness and sociability reexamined: A multicomponent analysis. *Journal of Personality and Social Psychology, 57*(5), 904-915. doi:10.1037/0022-3514.57.5.904

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75*(1), 219-229. doi:10.1037/0022-3514.75.1.219

Bushman, B. J., Baumeister, R. F., Thomaes, S., Ryu, E., Begeer, S., & West, S. G. (2009). Looking again, and harder, for a link between low self-esteem and aggression. *Journal of Personality, 77*(2), 427-446. doi:10.1111/j.1467-6494.2008.00553.x

Buss, A. H. (1980). *Self-consciousness and social anxiety.* San Fransisco, CA: Freeman.

Cacioppo, J. T., Glass, C. R., & Merluzzi, T. V. (1979). Self-statements and self-evaluation: A cognitive response analysis of heterosocial anxiety. *Cognitive Therapy and Research, 3*(3), 249-262. doi:10.1007/BF01185965

Cashin, W. E. (1999). Student ratings of teaching: Uses and misuses. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 25-44). San Fransisco, CA: Jossey-Bass.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*(5), 494-518.

Centra, J. A., & Gaubatz, N. B. (2000). Is there a gender bias in student evaluations of teaching? *The Journal of Higher Education, 71*(1), 17-33. doi:10.2307/2649280

Chacko, T. I. (1983). Student ratings of instruction: A function of grading standards. *Educational Research Quarterly, 8*(2), 19-25.

Cheek, J. M., & Briggs, S. R. (1990). Shyness as a personality trait. In W. R. Crozier (Ed.), *Shyness and embarrassment: Perspectives from social psychology* (pp. 315-337). New York, NY: Cambridge University Press.

Cheek, J. M., & Buss, A. H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology, 41*(2), 330-339. doi:10.1037/0022-3514.41.2.330

Chemers, M. M., Hu, L., & Garcia, B. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology, 93*(1), 55-64. doi:10.1037/0022-0663.93.1.55

Crocker, J., Karpinski, A., Quinn, D. M., & Chase, S. K. (2003). When grades determine self-worth: Consequence of contigent self-worth for male and female engineering and psychology majors. *Journal of Personality and Social Psychology, 85*(3), 507-516. doi:10.1037/0022-3514.85.3.507

Crozier, W. R. (Ed.). (1990). *Shyness and embarrassment: Perspectives from social psychology*. New York, NY: Cambridge University Press.

Crozier, W. R. (1995). Shyness and self-esteem in middle childhood. *British Journal of Educational Psychology, 65*(1), 85-95. doi:j.2044-8279.1995.tb01133.x

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development, 21*(1), 5-21.

Czeschlik, T., & Nurk, H. C. (1995). Shyness and sociability: Factor structure in a German sample. *European Journal of Psychological Assessment, 11*(2), 122-127. doi:10.1027/1015-5759.11.2.122

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*(11), 1198-1209. doi:10.1037/0003-066X.52.11.1198

Davis, M. H., & Stephan, W. G. (1980). Attribution of exam performance. *Journal of Applied Social Psychology, 10*(3), 235-248. doi:10.1111/j.1559-1816.1980.tb00706.x

Doyle, K. O. Jr. (1983). *Evaluating teaching.* Lexington, MA: Lexington Books.

Feeley, T. H. (2002). Evidence of halo effects in student evaluations of communication instruction. *Communication Education, 51*(3). 225-236. doi:10.1080/03634520216519

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education, 9*(3), 199-242. doi:10.1007/BF00976997

Feldman, K. A. (1992). College students' views of male and female college teachers: Part I – evidence from the social laboratory and experiments. *Research in Higher Education, 33*(3), 317-375. doi:10.1007/BF00992265

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II – evidence from students' evaluations of their classroom teachers. *Research in Higher Education, 34*(2), 151-211. doi:10.1007/BF00992161

Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 368-395). New York, NY: Agathon Press.

Felton, J., Mithcell, J., & Stinson, M. (2004). Web-based student evaluations of professors: The relations between perceived quality, easiness and sexiness. *Assessment and Evaluation in Higher Education, 29*(1), 91-108. doi:10.1080/0260293032000158180

Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness and other issues: Student evaluations of professors on ratemyprofessors.com. *Assessment and Evaluation in Higher Education, 33*(1). 45-61. doi: 10.1080/02602930601122803

Ford, D. H. (1987). *Humans as self-constructing living systems.* Hillsdale, NJ: Lawrence Erlbaum.

Franzoi, S. L. (1983). Self-concept differences as a function of private self-consciousness and social anxiety. *Journal of Research in Personality, 17*(3), 275-287. doi:10.1016/0092-6566(83)90019-3

Freng, S., & Webber, D. (2009). Turning up the heat on online teaching evaluations: Does "hotness" matter? *Teaching of Psychology, 36*(3), 189-193. doi:10.1080/00986280902959739

Gage, N. L. (1961). The appraisal of college teaching. *The Journal of Higher Education, 32*(1), 17-22. doi:10.2307/1980320

Gest, S. (1997). Behavioral inhibition: Stability and associations with adaptation from childhood to early adulthood. *Journal of Personality and Social Psychology, 72*(2), 467-475. doi:10.1037/0022-3514.72.2.467

Gigliotti, R. J., & Buchtel, F. S. (1990). Attributional bias and course evaluations. *Journal of Educational Psychology, 82*(2), 341-351. doi:10.1037/0022-0663.82.2.341

Goolsby, L. L., & Chaplin, W. F. (1988). The impact of normative, ipsative, and idiothetic information on feelings about academic performance. *Journal of Research in Personality, 22*(4), 445-464. doi:10.1016/0092-6566(88)90003-7

Gough, H. G., & Thorne, A. (1986). Positive, negative, and balanced shyness: Self-definitions and the reactions of others. In W. H. Jones, J. M. Cheek, & S. R. Briggs (Eds.), *Shyness: Perspectives on research and treatment* (pp. 205-225). New York, NY: Plenum Press.

Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student Course Evaluations: Research, Models and Trends.* Toronto: Higher Education Quality Council of Ontario.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*(11), 1209-1217. doi:10.1037/0003-006X.52.11.1209

Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology, 29*(4), 410-425. doi:10.1016/j.cedpsych.2003.11.001

Gurung, R. A., & Vespia, K. M. (2007). Looking good, teaching well? Linking liking, looks, and learning. *Teaching of Psychology, 34*(1), 5-10. doi:10.1080/00986280709336641

Heider, F. (1976). A conversation with Fritz Heider. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 47-61). Hillsdale, NJ: Erlbaum.

Heilman, J. D., & Armentrout, W. D. (1936). The rating of college teachers on ten traits by their students. *Journal of Educational Psychology, 27*(3), 197-216. doi:10.1037/h0054965

Henderson, L. (2002). Fearfulness predicts self-blame and shame in shyness. *Personality and Individual Differences, 32*(1), 79-93. doi:10.1016/S0191-8869(01)00007-1

Hildebrand, M., Wilson, R. C., & Dienst, E. R. (1971). *Evaluating university teaching.* Berkley, CA: University of California Center for Research and Development in Higher Education.

Hillman, K. (2005). The first-year experience: The transition from school to university and TAFE in Australia. *LSAY Research Reports.* Longitudinal surveys of Australian youth research report; n. 40. http://research.acer.edu.au/lsay_research/44

Isen, A. M., & Means, B. (1983). The influence of positive affect on decision-making strategy. *Social Cognition, 2*(1), 18-31. doi:10.1521/soco.1983.2.1.18

Izard, C. E., & Hyson, M. C. (1986). Shyness as a discrete emotion. In W. H. Jones, J. M. Cheek, & S. R. Briggs (Eds.), *Shyness: Perspectives on research and treatment* (pp. 147-160). New York, NY: Plenum Press.

Jones, W. H., Cheek, J. M., & Briggs, S. R. (Eds.). (1986). *Shyness: Perspectives on research and treatment.* New York, NY: Plenum Press.

Jones, E. E., Kannouse, D. E., Kelley, H. H., Nisbett, R. E., Valins, S., & Weiner, B., Eds. (1972). *Attribution: Perceiving the Causes of Behavior.* Morristown, NJ: General Learning Press.

Jones, D. N., & Paulhus, D. L. (2010). Different provocations trigger aggression in narcissists and psychopaths. *Social Psychological and Personality Science, 1*(1), 12-18. doi:10.1177/1948550609347591

Kelly, M. (2012). Student evaluations of teaching effectiveness: Considerations for Ontario universities. *Council of Ontario Universities Academic Colleagues Discussion Paper,* 1-15.

Kerr, A. (2011). *Teaching and Learning in Large Classes at Ontario Universities: An Exploratory Study.* Toronto: Higher Education Quality Council of Ontario.

Kierstead, D., D'Agostino, P., & Dill, H. (1988). Sex role stereotyping of college professors: Bias in students' ratings of instructors. *Journal of Educational Psychology, 80*(3), 342-344. doi:10.1037/0022-0663.80.3.342

Konrath, S., Bushman, B. J., & Campbell, W. K. (2006). Attenuating the link between threatened egotism and aggression. *Psychological Science, 17*(11), 995-1001. doi:10.1111/j.1467-9280.2006.01818.x

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review, 18*(1), 59-63. doi:10.1016/S0272-7757(98)00004-1

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. doi:10.1037/0022-3514.77.6.1121

Leary, M. R., Kowalski, R. M., & Campbell, C. D. (1988). Self-presentational concerns and social anxiety: The role of generalized impression expectancies. *Journal of Research in Personality, 22*(3), 308-321. doi:10.1016/0092-6566(88)90032-3

Lee, Y. T., & Seligman, M. E. (1997). Are Americans more optimistic than the Chinese? *Personality and Social Psychology Bulletin, 23*(1), 32-40. doi:10.1177/0146167297231004

Lovel, G. D., & Haner, C. F. (1955). Force-choice applied to college faculty rating. *Educational and Psychological Measurement, 15*(3), 291-304. doi:10.1177/001316445501500309

Magnusson, D., & Stattin, H. (2006). The person in context: A holistic-interactionistic approach. In W. Damon & R. M. Lerner (Eds.), *Handbook of Child Psychology. Vol. 1: Theoretical Models of Human Development* (pp. 400-464). New York, NY: Wiley.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal for Educational Research, 11*(3), 253-388. doi:10.1016/0883-0355(87)90001-2

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*(11), 1187-1197. doi:10.1037/0003-006X.52.11.1187

Martin, A. J., Marsh, H. W., & Debus, R. L. (2001). Self-handicapping and defensive pessimism: Exploring a model of predictors and outcomes from a self-protection perspective. *Journal of Educational Psychology, 93*(1), 87-102. doi:10.1037/0022-0663.93.1.87

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*(11), 1218-1225. doi:10.1037/0003-066X.52.11.1218

Meleshko, K. G., & Alden, L. E. (1993). Anxiety and self-disclosure: Toward a motivational model. *Journal of Personality and Social Psychology, 64*(6), 1000-1009. doi:10.1037/0022-3514.64.6.1000

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attribution? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*(5), 711-747. doi:10.1037/0033-2909.130.5.711

Neufeld, J. E., Rasmussen, H. N., Lopez, S. J., Ryder, J. A., & Magyar-Moe, J. L. (2006). The engagement model of person-environment interaction. *The Counseling Psychology, 34*(2), 1-16. doi:10.1177/0011000005281319

O'Banion, K., & Arkowitz, H. (1977). Social anxiety and selective memory for affective information about the self. *Social Behavior and Personality, 5*(2), 321-328. doi:10.2224/sbp.1977.5.2.321

Olivares, O. J. (2011). Student interest, grading leniency, and teacher ratings: A conceptual analysis. *Contemporary Educational Psychology, 26*(3), 382-399. doi:10.1006/ceps.2000.1070

Papps, B. P., & O'Carroll, R. E. (1998). Extremes of self-esteem and narcissism and the experience of anger and aggression. *Aggressive Behavior, 24*(6), 421-438. doi:10.1002/(SICI)1098-2337(1998)24:g<421::AID-AB3>3.0.CO;2-3

Patrick, C. L. (2011). Student evaluations of teaching: Effects of the big five personality traits, grades and the validity hypothesis. *Assessment and Evaluation in Higher Education, 36*(2), 239-249. doi:10.1080/02602930903308258

Pausch, R. (2007). Last Lecture: Really Achieving your Childhood Dreams. Carnegie Mellon University, Pittsburgh, PA.

Peterson, C., & Barrett, L. C. (1987). Explanatory style and academic performance among university freshmen. *Journal of Personality and Social Psychology, 53*(3), 603-607. doi:10.1037/0022-3514.53.3.603

Pohlman, J. T. (1977). A multivariate analysis of selected class characteristics and student ratings of instruction. *Multivariate Behavioral Research, 10*(1), 81-91. doi:10.1207/s15327906mbr1001_5

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment and Evaluation in Higher Education, 33*(2), 143-154. doi:10.1080/02602930601127869

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*(2), 590. doi:10.2466/pr0.1979.45.2.590

Raskin, R. N., & Hall, C. S. (1981). The Narcissistic Personality Inventory: alternate form reliability and further evidence of construct validity. *Journal of Personality Assessment, 45*(2), 159-162. doi:10.1207/s15327752jpa4502_10

Raskin, R., & Terry, H. (1988). A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*(5), 890-902. doi:10.1037/0022-3514.54.5.890

Rizley, R. (1978). Depression and distortion in the attribution of causality. *Journal of Abnormal Psychology, 54*(5), 890-902. doi:10.1037/0022-3514.54.5.890

Rocca, K. A., & McCroskey, J. C. (1999). The interrelationship of student ratings of instructors' immediacy, verbal aggressiveness, homophily, and interpersonal attraction. *Communication Education, 48*(4), 308-316. doi:10.1080/03634529909379181

Ross, M., & Fletcher, G. J. (1985). Attribution and social perception. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 73-122). New York, NY: Random House.

Saroyan, A., & Amundsen, C. (2001). Evaluating university teaching: Time to take stock. *Assessment & Evaluation in Higher Education, 26*(4), 341-353. doi:10.1080/02602930120063493

Scherr, F., & Scherr, S. S. (1990). Bias in student evaluations of teacher effectiveness. *Journal of Education for Business, 65*(8), 356-358.

Schlenker, B. R., & Leary, M. (1982). Social anxiety and self-presentation: A conceptualization and a model. *Psychological Bulletin, 92*(3), 641-669. doi:10.1037/0033-2909.92.3.641

Seldin, P. (1998). How colleges evaluate teaching: 1988 vs. 1998. *American Association of Higher Education Bulletin, 50*(7), 3-7.

Seligman, M. E., Nolen-Hoeksema, S., Thornton, N., & Thornton, K. M. (1990). Explanatory style as a mechanism of disappointing athletic performance. *Psychological Science, 1*(2), 143-146. doi:10.1111/j.1467-9280.1990.tb00084.x

Shepperd, J. A., Arkin, R. M., & Slaughter, J. (1995). Constraints on excuse making: The deterring effects of shyness and anticipated retest. *Personality and Social Psychology Bulletin, 21*(10), 1061-1072. doi:10.1177/01461672952110007

Shepperd, J. A., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bas. *Social and Personality Psychology Compass, 2*(2), 895-908. doi:10.1111/j.1751-9004.2008.00078.x

Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment and Evaluation in Higher Education, 25*(4), 397-405. doi:10.1080/713611436

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education, 22*(3), 199-213. doi:10.1007/BF02620508

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin, 26*(11), 1329-1342. doi:10.1177/0146167200263002

Smith, R. E., & Sarason, I. G. (1975). Social anxiety and the evaluation of negative interpersonal feedback. *Journal of Consulting and Clinical Psychology, 43*(3), 429. doi:10.1037/h0076855

Snyder, C. R., & Clair, M. (1976). Effects of expected and obtained grades on teacher evaluation and attribution of performance. *Journal of Educational Psychology, 68*(1), 75-82. doi:10.1037/0022-0663.68.1.75

Stucke, T. S., & Sporer, S. L. (2002). When a grandiose self-image is threatened: Narcissism and self-concept clarity as predictors of negative emotions and aggression following ego-threat. *Journal of Personality, 70*(4), 509-523.

Sweeney, P. D., Anderson, K., & Bailey, S. (1986). Attributional style in depression: A meta-analytic review. *Journal of Personality and Social Psychology, 50*(5), 974-991. doi:10.1037/0022-3514.50.5.974

Tam, M. (2001). Measuring quality and performance in higher education. *Quality in Higher Education, 7*(1), 47-54. doi:10.1080/1358320120045076

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A., Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review, 107*(3), 411-429. doi:10.1037/0033-295X.107.3.411

Teglasi, S. E., & Hoffman, M. A. (1982). Causal attributions of shy subjects. *Journal of Research in Personality, 16*(3), 376-385. doi:10.1016/0092-6566(82)90033-2

Twenge, J. M., & Campbell, W. K. (2003). "Isn't if fun to get the respect we're going to deserve?" Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin, 29*(2), 261-272. doi:10.1177/0146167202239051

Vaillancourt, T. (2005). Indirect aggression among humans: Social construct or evolutionary adaptation? In R. E. Tremblay, W. H. Hartup, & J. Archer (Eds.), *Developmental origins of aggression.* New York, NY: Guilford Press.

Vaillancourt, T. (2013). Students aggress against professors in reaction to receiving poor grades: An effect moderated by student narcissism and self-esteem. *Aggressive Behavior, 39(1*), 71-84. doi:10.1002/ab.21450

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education, 23*(2), 191-211. doi:10.1080/0260293980230207

Watson, P. J., Grisham, S. O., Trotter, M. V., & Biderman, M. D. (1984). Narcissism and empathy: Validity evidence for the narcissistic personality inventory. *Journal of Personality Assessment, 48*(3), 301-305. doi:10.1207/s15327752jpa4803_12

Weiner, B. (1974). *Achievement motivation and attribution theory.* Morristown, NJ: General Learning Press.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer-Verlag.

Wolfer, T. A., & Johnson, M. N. (2003). Re-evaluating student evaluation of teaching: The teaching evaluation form. *Journal of Social Work Education, 39*(1), 111-121.

Woodhouse, D. (1998). Auditing research and the research/teaching nexus. *New Zealand Journal of Educational Studies, 33*(1), 39-53.

Worthington, A. G., & Wong, P. T. (1979). Effects of earned and assigned grades on student evaluations of an instructor. *Journal of Educational Psychology, 71*(6), 764-775. doi: 10.1037/0022-0663.71.6.764

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education, 12*(1), 55-76. doi: 10.1080/13562510601102131

Zimbardo, P. G. (1977). *Shyness: What it is, what to do about it.* Menlo Park, CA: Addison-Wesley.

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality, 47*(2), 248-287. doi: 10.1111/j.1467-6494.1979.tb00202.x

Table 1

Frequencies of Content Analysis Coding Categories: Attribution

		Frequency	Percent
Positive Condition	External	76	26.2
	Internal	13	4.5
	Neutral	201	69.3
Negative Condition	External	123	45.1
	Internal	15	5.5
	Neutral	135	49.5
Total	External	199	35.3
	Internal	28	5.0
	Neutral	336	59.7

Table 2

Frequencies of Content Analysis Coding Categories: Negative Condition

		Frequency	Percent
Time	No	162	59.3
	Yes	111	40.7
	Total	273	100.0
Bias	No	263	96.3
	Yes	10	3.7
	Total	273	100.0
Revenge	No	268	98.2
	Yes	5	1.8
	Total	273	100.0
Questions Competence	No	250	91.6
	Yes	23	8.4
	Total	273	100.0
Rushed	No	246	90.1
	Yes	27	9.9
	Total	273	100.0
Bad Writer	No	264	96.7
	Yes	9	3.3
	Total	273	100.0
Competent Evaluator	No	206	75.5
	Yes	67	24.5
	Total	273	100.0
Lack of Detail/Vague	No	62	22.7
	Yes	211	77.3
	Total	273	100.0
No Resources	No	237	86.8
	Yes	36	13.2
	Total	273	100.0
Unfair Marking	No	243	89.0
	Yes	30	11.0
	Total	273	100.0
Confident Writer	No	270	98.9
	Yes	3	1.1
	Total	273	100.0
Helpful Feedback	No	265	97.1
	Yes	8	2.9
	Total	273	100.0

Table 3

Frequencies of Content Analysis Coding Categories: Positive Condition

		Frequency	Percent
Time	No	224	77.2
	Yes	66	22.8
	Total	290	100.0
Bias	No	284	97.9
	Yes	6	2.1
	Total	290	100.0
Revenge	No	290	100.0
	Yes	0	0
	Total	290	100.0
Questions Competence	No	228	78.6
	Yes	62	21.4
	Total	290	100.0
Rushed	No	268	92.4
	Yes	22	7.6
	Total	290	100.0
Bad Writer	No	283	97.6
	Yes	7	2.4
	Total	290	100.0
Competent Evaluator	No	164	56.6
	Yes	126	43.4
	Total	290	100.0
Lack of Detail/Vague	No	116	40.0
	Yes	174	60.0
	Total	290	100.0
No Resources	No	282	97.2
	Yes	8	2.8
	Total	290	100.0
Unfair Marking	No	285	98.3
	Yes	5	1.7
	Total	290	100.0
Writer Confident	No	285	98.3
	Yes	5	1.7
	Total	290	100.0
Helpful Feedback	No	265	91.4
	Yes	25	8.6
	Total	290	100.0

Table 4

Summary of Chi-Square Frequencies

		Negative	Positive
External	Count	123	76
	% of Total	21.8	13.5
Internal	Count	15	13
	% of Total	2.7	2.3
Neutral	Count	135	201
	% of Total	24.0	35.7