Measuring Education Science Students‘ Statistics Anxiety
Conceptual Framework, Methodological Considerations, and Empirical Analyses

Research Report
Leibniz University Hannover: Institute of Psychology 2018

Dr. Günter Faber
Leibniz University Hannover, Institute of Psychology

Dr. Heike Drexler
Leibniz University Hannover, Institute of Psychology

MA Alexander Stappert
University of Vechta, Department of Social and Education Sciences

BA Joana Eichhorn
Leibniz University Hannover, Education Science Master’s Programme
Measuring education students’ statistics anxiety.
Conceptual framework, methodological considerations, and empirical analyses.
Research report

Günter Faber
Heike Drexler
Alexander Stappert
Joana Eichhorn

Leibniz University Hannover
Faculty of Humanities
Institute of Psychology
Copyright © Faber, Drexler, Stappert, Eichhorn 2018
Table of contents

1 Abstract .......................................................................................................................... 2
2 Introduction .................................................................................................................. 3
  2.1 Conceptual perspectives on statistics anxiety ....................................................... 4
  2.2 Methodological perspectives on statistics anxiety measures ......................... 8
  2.3 Approaching refined measurement ................................................................. 9
3 Validation framework and objectives ...................................................................... 11
4 Method
  4.1 Participants and procedure ............................................................................. 15
  4.2 Measures ......................................................................................................... 16
  4.3 Data analyses ................................................................................................. 18
5 Results
  5.1 Scale formation ............................................................................................... 18
  5.2 Validation results ............................................................................................ 21
  5.3 Additional multivariate analyses .................................................................... 23
6 Conclusion ............................................................................................................... 25
7 References .............................................................................................................. 29
8 Appendix ............................................................................................................... 46
1 Abstract

At higher education levels, the acquisition and application of research methods appears to represent a great challenge for a considerable number of students. As relevant empirical findings consistently demonstrated, many students in social or human science programmes tend to evolve and maintain negative attitudes against compulsory method courses and, in particular, perceive statistical demands as threatening study events – thus, being at risk of developing a heightened level of statistics anxiety. Therefore, university settings need to pay attention to this phenomenon and strive for a sound analysis of the students’ concern. Accordingly, over the past decades several questionnaires had been developed to assess university learners’ statistics anxiety in order to get appropriate information for conceptualizing most adaptive instructional and guidance strategies. However, against theoretical and methodological background of test anxiety research, current instruments for assessing university students’ statistics anxiety prevailing emphasize the affective construct component. In order to unfold the construct in a more exhaustive and differentiated manner, a scale for measuring university students’ worry, avoidance, and emotionality cognitions was developed. In two samples of education science majors the present study aimed at analyzing the scale’s psychometric properties and at gaining preliminary validation results. In both samples, principal component analysis led to the formation of a unidimensional scale which appeared to be sufficiently reliable. Its relations to domain-specific self-belief and background variables turned out as theoretically expected – thus, for the time being the scale should claim criterion validity. In particular, the scale’s total sum score was demonstrated to substantially correlate with the students’ mathematical self-concept, entity beliefs, and negative instrumental values.

Keywords: statistics anxiety, math self-concept, implicit entity beliefs, utility values
2 Introduction

Higher education settings usually need to manage a broad range of students’ individual background, learning approaches, and study outcomes. Therefore, in order to ensure a most adaptive and effective academic environment, they must rely on appropriate conceptualizations of the students’ learning processes which should help to analyze the individual and contextual factors of university achievement and to explain existing differences in the students’ performance – and, eventually, to access reliable and valid information about the substantial determinants of university success or failure. For that purpose, educational research had offered several modeling perspectives to unfold the complexity of learning processes being typical for university contexts. Though each of them focusing on learning and achievement in a different manner, all these conceptualizations widely agree in stressing the importance of students’ individual aptitudes, body of knowledge, and motivational characteristics (Credé & Kunzel, 2008; De Clercq, Galand, Dupont, & Frenay, 2013; Entwistle, 2007; Price, 2014; van Rooij, Brouwer, Bruinsma, Jansen, Donche, & Noyens, 2018). For the most part, empirical research in the field had verified and further differentiated these conceptual assumptions. In particular, relevant meta-analyses unanimously demonstrated the crucial role of students’ cognitive abilities, prior knowledge, and latest school grades to predict academic outcomes at university level. Furthermore, they evidenced the impact of students’ motivational orientations on their academic attainment (Richardson, Abraham, & Bond, 2012; Robbins, Lauver, Le, Davis, Langley, & Carlstrom, 2004; Schneider & Preckel, 2017).

Especially, the motivational orientations cover various types and levels of academic self-beliefs which essentially may regulate the students’ learning approach and, thus, affect their academic outcomes (Pintrich, 2004). The students’ competence and control beliefs must be recognized as cognitive-motivational key constructs (Krampen, 1988; Schunk & Zimmerman, 2006). Competence beliefs refer to an individual’s perceived capabilities to succeed in a given academic activity or task (Eccles & Wigfield, 2002). Control beliefs refer to an individual’s perceived likelihood of accomplishing desired academic outcomes or of avoiding undesired academic outcomes by means of his or her own behavioral attempts (Perry, Hall, & Ruthig, 2005). Both types of beliefs result from an individual’s cumulative success or failure experiences and must be seen as important personal resources to cope with academic requirements in a certain educational setting. Over time, competence and control beliefs will operate in a mutually reinforcing manner and contribute to implement the students’ more or less adaptive learning approach (Drew & Watkins, 1998; Robbins, Lauver, De, Davis, Langley, & Carlstrom, 2004; van Overvalle, 1989). Accordingly, students with high competence and control beliefs will develop a strong sense of self-efficacy and evolve learning strategies to persistently endeavor successful outcomes – whereas students with low competence and control beliefs will develop strong concerns not to master the academic tasks at hand and possibly rely on less adaptive learning strategies. In the latter case, students are seriously prone to increasingly experience feelings of personal threat and, in the long term, to stabilize trait-like test anxieties (Putwain, 2018; Respondek, Seufert, Stupnisky, & Nett, 2017). As empirical findings from various university settings consistently substantiated, an individually heightened level of test anxiety predicts less useful learning strategies (Spada, Nikcevic, Moneta, & Ireson, 2006; Virtanen, Nevgi, & Niemi, 2013) and leads to significantly lowered learning outcomes (Credé &
Kunzel, 2008; Feinberg & Halperin, 1978; Hancock, 2001; Kassim, Hanafi, & Hancock, 2007; Schneider & Preckel, 2017). Hence, the emergence, maintenance, and consequences of sustained test anxieties indicate a potential risk factor of the students’ learning – which foreseeably threatens to impede their academic perspectives in a certain study domain or programme. Consequently, higher education settings should carefully pay attention to this phenomenon and strive for the implementation of appropriate instructional or guidance strategies.

With all that, the learning of research methods, most notably the acquisition of statistical knowledge and competencies, appears to represent a stress or anxiety provoking domain in higher education contexts. Quite a number of undergraduate and graduate students of social sciences, education, psychology and business programmes struggles with statistics (Mji & Onwuegbuzie, 2004; Onwuegbuzie & Wilson, 2003). Moreover, in comparison with other academic courses their individually reported level of anxiety appears to be highest in the statistics domain (Dykeman, 2011). When dealing with the requirements of quantitative method courses which commonly are compulsory for earning their degree, these students mostly suffer from strong failure expectations and frequently experience feelings of apprehension and personal threat. As well, they perceive methodological competencies difficult to acquire and less useful for current studies and later professional development (Murtonen & Lehtinen, 2003; Murtonen, Olkinuora, Tynjälä, & Lehtinen, 2008; Zeidner, 1991). As a result, they are at risk for developing and maintaining a heightened level of anxiety in the face of statistical analyses – in particular, when being confronted with statistical tasks of data gathering, processing, and interpreting (Cruise, Cash, & Bolton, 1985). Over the past decades, numerous empirical analyses from most diverse conceptual, methodological, and institutional backgrounds had yielded a vast amount of insights into the development, the structure, and the consequences of higher education students’ statistics anxiety – and, thus, advancingly established a broad-based research line to clarify and elaborate the construct.

2.1 Conceptual perspectives on statistics anxiety

Structurally, the students’ emerging statistics anxiety has to be considered a multidimensional construct reflecting the complex interplay of several cognitive, motivational, and physiological components (Rost & Schermer, 1989a). Against the background of relevant empirical findings from test anxiety research it can be defined as a domain-specific form of performance or evaluation anxiety which manifests as repeatedly occurring worry cognitions, task-irrelevant and interfering thoughts, marked states of emotional tension and physiological arousal (Zeidner, 1991). The worry component of test anxiety refers to the students’ mental anticipation of failure and its negative consequences, whereas the emotionality component refers to their feelings of tenseness, nervousness or distress, and the physiological component refers to their perceptions of bodily symptoms. As in some conceptual approaches the emotionality component includes physiological reactions, other conceptual approaches differentiate or extend these components (Cassady & Johnson, 2002; Schwarzer & Quast, 1985; Zeidner, 1998). Over the past decades, across a wide range of student samples and measurements there has been ample evidence for these components being distinguishable but mutually reinforcing (Deffenbacher, 1980; Hodapp & Benson, 1997; Kieffer & Reese, 2009; Liebert & Morris,
In most cases, they had been demonstrated to negatively affect the students’ learning processes and achievement outcomes. However, the worry component generally turned out to most strongly predict academic performance or test results (Cassady & Johnson, 2002; Deffenbacher, 1980; McIlroy, Bunting, & Adamson, 2000; Rana & Mahmood, 2010; Schwarzer & Jerusalem, 1992; Seipp, 1991; Steimmayr, Crede, McElvany, & Wirthwein, 2016; von der Embse, Jester, Roy, & Post, 2018). The debilitating effect of worry cognitions on students’ performance appears to be mainly caused by their strongly biased and task-irrelevant mode of information processing. High-anxious students usually tend to perceive threatening cues within a certain evaluation setting in an extremely attentive or even sensitive manner (Wine, 1982; Fernández-Castillo & Caurel, 2015; Putwain, Connors, & Symes, 2010). They produce more and stronger self-doubts which interfere with the process of task completion—in particular, by thinking more frequently and intensely about the inevitable failure they expect (Sarason, 1984; Schwarzer, 1996). Thus, high-anxious students’ attention and cognitions operate more self-focused than task-focused. As a result, their cognitive resources are to be seen heavily preoccupied and notably impeded (Ashcraft & Moore, 2009; Hopko, Ashcraft, Gute, Ruggiero, & Lewis, 1998; Richards, French, Keogh, & Carter, 2000).

In the long term, high-anxious students tend to use less adaptational coping strategies. They prefer behavioral responses to immediately reduce their cognitive worries and emotional apprehension—rather than basically enhancing their task-relevant knowledge and learning approach (Rost & Schermer, 1989b; Zeidner, 1998). However, these strategies neither will contribute to handle their feelings of threat and distress, nor will they facilitate their mastery of the academic task. Instead they will most likely lead to the failure outcome the students already fear for. From a transactional perspective, the emergence and maintenance of test anxiety crucially contributes to establishing a self-reinforcing circle of negative anxiety-outcome relations (Spielberger & Vagg, 1995; Zeidner, 1998).

Based on this conclusive body of evidence a theoretically and methodologically sound framework for the statistics anxiety construct should implicitly take into account its cognitive, emotional, and physiological components. In particular, it is essential that it addresses the issue of relevant worry cognitions as it is assumed that they are closely linked to the debilitating effects of statistics anxiety on statistical learning and performance.

To date, however, most empirical analyses in the field have emphasized the emotional and physiological components of statistics anxiety (Cruise, Cash, & Bolton, 1985). They define the construct as feelings of anxiety or as habitual anxiety in the face of statistically loaded situations (Macher, Paechter, Papousek, & Ruggeri, 2012; Onwuegbuzie & Wilson, 2003). Admittedly, the measurement items used in these empirical analyses cover a wide range of statistical tasks that students typically encounter in everyday study or the context of their course. These items can be empirically assigned to various situation- or task-specific dimensions.

Thus, factor analyses of the task- and course-related items of Zeidner’s (1991) Statistics Anxiety Inventory led to the development of a content- and a test-specific subscale. Likewise, factor analyses of the widely used Statistics Anxiety Rating Scale (Cruise, Cash, & Bolton, 1985) and the conceptually related Statistical Anxiety Scale (Vigil-Colet, Lorenzo-Seva, & Condon, 2008) provided separate subcomponents concerning the students’ interpretation and

Similarly, the Statistics Anxiety Measure (Earp, 2007; Vahedi, Farrokhi, & Bevrani, 2011) and the Statistics Comprehensive Anxiety Response Evaluation (Griffith, Mathna, Sappington, Turner, Evans, Gu, Adams, & Morin, 2014) revealed some distinct task- or situation-specific subcomponents referring to statistically relevant course requirements and situations. In summary, these approaches to measuring the construct of statistics anxiety undoubtedly represent typically anxiety-evoking task features and test situations in a most elaborate way. However, with the exception of the Statistics Anxiety Measure (Earp, 2007) and the short research scale Hong and Karstensson (2002) used – both of which include some worry items – it is notable that all the other instruments fail to integrate students’ cognitive, emotional, and physiological anxiety reactions, in particular with regard to the critical worry component.

Williams’s (2013, 2015) findings provide support to this conceptual shortcoming, as they demonstrated that scores on the interpretation and test/class anxiety subscales of the Statistics Anxiety Rating Scale (Cruise, Cash, & Bolton, 1985) were only moderately correlated with students’ tendency to worry.

In contrast, a concurrently operating research line in the test anxiety field had already decomposed the statistics anxiety construct and assessed the students’ worry and emotionality responses separately. However, in most cases a composite score including both components was used because of both components’ high interrelation (Benson, 1989; Benson, Bandalos, & Hutchinson, 1994; Finney & Schraw, 2003; González, Rodrígez, Fafilde, & Carrera, 2016; Hong & Carstenssson, 2002). That way, albeit merely having total anxiety scores available, the interpretation of students’ responses explicitly allowed for a traceable cognitive perspective. Furthermore, one relevant study had analyzed the relations of both the worry and emotionality component with several cognitive-motivational and achievement variables. According to the findings of test anxiety research, its results demonstrated the worry component to better explain the students’ performance in a statistics exam (Bandalos, Yates, & Thorndike-Christ, 1995).

As this research line essentially contributes to refine the statistics anxiety construct with respect to its motivationally operating components, its task- or situation-specific references appear less differentiated. That is, in all studies the items for assessing statistics anxiety referred exclusively to the taking of a statistical test or exam. This contextual limitation, hence, should challenge the representativity or content validity of the worry and emotionality measures, because their scores account only for a particular part of the relevant learning setting (Cohen, Manion, & Morrison, 2011; Haynes, Richard, & Kubany, 1995).

A most appropriate framework of the statistics anxiety construct should merge the conceptual strengths of both research lines and, therefore, consider the relevant worry, emotionality, and bodily reactions students will experience when dealing with the typically occurring task
features and examination procedures of their statistics course or of any other course – if necessarily requiring statistical knowledge and skills. Refining the construct that way, research in the field should advance in developing measurement tools which will help to determine the students’ statistics anxiety in a psychologically as well as contextually more differentiated and, thus, in a more representative manner. Consequently, their construction should thoroughly specify the construct’s key facets (Guttman & Greenbaum, 1998) and transform its a priori assumptions into most concrete operationalizations mapping out both relevant anxiety reactions and typically threatening situations (Zeidner, 1998).

Moreover, another issue crucial to the conceptualization of the statistics anxiety construct refers to the role of the students’ avoidance tendencies. Already in the very first beginning of empirical test anxiety research, Mandler and Sarason (1952) posited anxiety responses to manifest as “implicit attempts at leaving the test situation” (166). Subsequently, empirical findings provided support for this assumption and yielded sound evidence for students’ avoidance cognitions being substantially related to their anxiety responses (Blankenstein, Flett, & Watson, 1992; Galassi, Frierson, & Sharer, 1981; Genc, 2017; Hagtvet & Benson, 1997; Middleton & Midgley, 1997; Skaalvik, 1997; Zeidner, 1998). In particular, the analyses of Elliot and McGregor (1999), Pekrun, Elliot and Maier (2009), and Putwain and Symes (2012) demonstrated clearly that students with heightened avoidance orientations reported a higher extent of worry cognitions and lower scores on subsequent exam performance. Conceptually, these empirical findings primarily originate from achievement motivation research, in most cases from goal orientation theories which had developed several instruments to assess the students’ striving no to perform worse than others in a certain academic setting (Elliot & Murayama, 2008). Though not unfolding the avoidance issue in an exhaustive manner, this motivation approach should unarguably contribute to broadening and refining the test anxiety construct – and, thus, offer promising perspectives to integrate the role of students’ avoidance tendencies (Putwain, 2008).

Heretofore, only few conceptualizations of the test anxiety construct had addressed this issue and claimed the students’ escape or avoidance cognitions to constitute an essential part of their worries and to represent an important factor to elicit interfering, task-irrelevant thoughts (Pekrun, Goetz, Kramer, Hochstadt, & Molfenter, 2004; Schwarzer & Quast, 1985). Future research efforts are required to further determine the relevance of avoidance thoughts within the worry component. Accordingly, in order to strive for a most differentiated analysis of the construct, research in the field should benefit from an enhanced emphasis on students’ avoidance cognitions. Hence, it should provide appropriate measurements being designed not only to assess the students’ worries about threatening failure outcomes, but also to inquire their thoughts to preferably avoid getting involved with threatening tasks or situations. Currently available questionnaires either do not consider the students’ avoidance cognitions at all (Griffith, Mathna, Sappington, Turner, Evans, Gu, Adams, & Morin, 2014; Onwuegbuzie & Wilson, 2003; Vigil-Colet, Lorenzo-Seva, & Condon, 2008) or include just a single item assessing avoidance behavior (Earp, 2007). They do not allow for a proper clarification of their role within the students’ worry responses. The same applies to one single study which had examined the students’ avoidance goals as predictors of statistics anxiety and, thus, demonstrated a moderate association (Zare, Rastegar, & Hosseini, 2011).
Research in the field should refine its conceptualization of students’ worry responses and develop instruments that explicitly capture avoidance cognitions with respect to statistically loaded tasks and situations. That way, an important step to refine the substantive and structural stage of construct validation would be done (Benson, 1998).

2.2 Methodological perspectives on statistics anxiety measures

Students’ statistics anxiety is commonly assessed using questionnaires consisting of items in Likert-scale format. With respect to statistically relevant task features or situations, they usually offer response categories which ask the students for each task- or situation-related item to rate the subjectively experienced intensity of being anxious. The response alternatives range from “no anxiety” to “high anxiety” in the Statistics Anxiety Rating Scale (Cruise, Cash, & Bolton, 1985), likewise from “not at all anxiety evoking” to “extremely anxiety evoking” in the Statistics Anxiety Inventory (Zeidner, 1991), from “not anxious” to “very anxious” in the Statistics Anxiety Measure (Earp, 2007), from “no anxiety” to “considerable anxiety” in the Statistical Anxiety Scale (Vigil-Colet, Lorenzo-Seva, & Condon, 2008), and from “no anxiety” to “severe anxiety” in the Statistics Comprehensive Anxiety Response Evaluation (Griffith, Mathna, Sappington, Turner, Evans, Gu, Adams, & Morin, 2014).

Thereby, it is left up to the student to interpret the meaning of these categories and to decide implicitly, to what extent the answer would reflect her or his cognitive worries, negative emotions, or bodily reactions. Although these anxiety components have been demonstrated to be substantially associated, the ambiguity of anchor labels (Podsakoff, Mackenzie, Lee, & Podsakoff, 2003) should hinder a most differentiated and valid clarification of the students’ anxious reactions and, hence, should not sufficiently comply with the conceptual perspectives on, and the empirical knowledge about, the test anxiety construct (Rost & Schermer, 1989a). Rather, these response categories appear to be principally at risk for compressing or even covering the students’ actual understanding of the response she or he has individually decided for (Krosnick & Presser, 2010). As a consequence, this response format potentially might contribute to establish a method-driven restriction in the interpretation of research data. Therefore, from both a conceptual and methodological perspective for future scale development in the field, it is recommended to apply response labels asking for the students’ agreement with clearly and distinguishably worded items to assess their cognitive, emotional, and physiological reactions (Cohen, Manion, & Morrison, 2011; Rost & Schermer, 1989a).

Furthermore, the anchors currently in use might be considered emotionally loaded and hence to have the potential to trigger biased responses (Lee, 2006; Podsakoff, Mackenzie, Lee, & Podsakoff, 2003). Principally, this biased responses will operate in different ways and produce contradictory effects: on the one hand, those students experiencing a heightened level of statistics anxiety should be prone to avoid extreme statements such as “high anxiety”, “extremely anxiety evoking”, or “considerable anxiety” and, thus, tend to report a socially and emotionally more acceptable extent of statistics anxiety (Lunneborg, 1964; Zeidner, 1998) – inasmuch as they would judge their anxious feelings as manifestation of personal weakness. Therefore, they should be motivated to report a somewhat lowered or trivialized level of anxiety and, thus, to protect their sense of self-worth (Harter, 1990). They might also feel embarrassed by the response wording and try to repress their anxiety symptoms. Hence, the response wording
would provoke a certain kind of coping strategy (Rost & Schermer, 1989b). On the other hand, students with a heightened level of statistics anxiety should process these anchor labels in a most sensitive way as threatening stimuli (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & Ijzendoorn, 2007; Fernández-Castillo & Caurel, 2015; Lawson, 2006) and, accordingly, report an inflated level of statistics anxiety (Mathews & MacLeod, 2002).

Although there is only scarce empirical evidence to reasonably clarify the prevalence and direction of that emotional response bias, no matter what, its potentially confounding effect on research data must be seen as most plausible (Gaye-Valentine & Credé, 2013). In particular, possible response bias should affect the students’ responses in an uncontrollable way and seriously diminish their validity, eventually. Work on developing new measurement instruments in the field should take this methodological flaw into account and use unloaded response categories as a precaution.

2.3 Approaching refined measurement

To overcome the conceptual limitations and methodological shortcomings of current instruments as a very first attempt a new scale to assess university students’ statistics anxiety was developed. Nevertheless, it should adopt the particular strengths of existing instruments. Thus, it was assigned to approach a refined measure of the construct by meeting the following criteria: Its items should (1) specifically take into account the students’ anxiety responses in a most differentiable way and, hence, consider their worry and avoidance cognitions as well as their emotional reactions. Thereby, its items should embed the various anxiety reactions (2) into a representative range of statistically loaded task features and course situations the students would typically encounter. In all this, its items should use (3) statements and response labels which do not offer emotionally or motivationally biasing wording.

The construction of this scale for measuring university students’ “Worry, Avoidance, and Emotionality Cognitions Encountering Statistical Demands” (WAESTA) largely followed the procedure of facet theory using a mapping sentence (Guttman & Greenbaum, 1998). This mapping sentence served as a heuristic device to cover all major facets of the construct and, thus, to achieve sufficient content validity (Edmundson, Koch, & Silverman, 1993). A pool of eligible items was drawn up using this constitutive mapping sentence which included conceptually relevant anxiety components, situational references, and intended response categories (Zeidner, 1998).

In particular, each item to represent the statistics anxiety domain was specified with respect to four key facets: a relevant reaction facet referring to the worry, avoidance, and emotionality component, and three contextual facets referring to the (1) outcome in a statistics exam, the (2) individual learning of statistical procedures and handling of statistical demands, and the (3) public mastering of statistical content. Furthermore, an additional range facet defined the response categories to assess the students’ perceived magnitude of individual anxiety reactions. As seemingly appropriate response range a four-point format was decided – in order to avoid artificial complexities in the respondents’ decision making (Lee & Paek, 2014; Lozano, Garcia-Cueto, & Muñiz, 2008) but instead to ensure a cognitive-motivationally realistic as well as just manageable number of rating references.

For instance, statistics anxiety should be more or less manifest as a student’s experienced or
anticipated worries about the anticipated exam outcome, the understanding of a certain statistical formula and the presentation of research findings.

This four-facet mapping sentence was used as a cross-classification template to systematically operationalize the statistics anxiety construct as it allowed to combine the various elements of facets in a most differentiated manner (Hox, 1997). That way, a final scale version with 17 four-point Likert-type rating items was built (Table 1).

All items concerned a mentally imaginable situation the students should easily manage to anticipate. Eight items referred to the students’ worries about their potentially expected failure to master the course exam and to cope with several statistical requirements. Four items concerned their cognitions to preferably avoid the statistics course and particular statistical demands. Five items are related to their emotional tension when being confronted with a certain statistical task (Appendix).

Table 1. WAESTA items assessing relevant anxiety reactions to statistically loaded task features and course situations

<table>
<thead>
<tr>
<th>Test Anxiety Component</th>
<th>Course Exam Outcome</th>
<th>Understanding Explanation Application</th>
<th>Oral Task Presentation Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worry</td>
<td>01, 14, 16</td>
<td>05, 08</td>
<td>10, 11, 12</td>
</tr>
<tr>
<td>Avoidance</td>
<td>03</td>
<td>06, 17</td>
<td>04</td>
</tr>
<tr>
<td>Emotionality</td>
<td>02, 07, 09, 13</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

Example (Item 07): _quite nervous_ explain a chart

“I would be _quite nervous_ if I were asked to _explain a chart_ from a research report.”

Response range Does not apply (1) … … Applies in full (4)

As statistically indicated task requirements for the understanding of course contents, the interpretation of quantitative research results as well as the application of statistical formulas and procedures were considered. Likewise, the oral presentation and explanation of statistical content in the public course situation was included. To warrant conceptual clarity, the avoidance items should neither tap the students’ avoidance reactions by suppressing or substituting individually occurring threat cognitions (Williams, 2015), nor should they refer to the students’ actual avoiding strategies to cope with disliked or threatening academic events (Onwuegbuzie, 2004; Rost & Schermer, 1989b). Rather they should operationalize the students’ mentally processed avoidance thoughts or even escape illusions before or during statistical task completion – and, thus, might indicate a specific subcomponent of worry cognitions. Moreover, the component of physiological symptoms or bodily tensions was not explicitly addressed. Instead it was thought to be indirectly inferred from the emotionality items. Conceptually, this restriction seemed to be justifiable, since the students’ perceived affective state should always reflect their actual physiological arousal (Zeidner, 1998).
3 Validation framework and objectives

As the WAESTA scale should be examined for the first time, the present study was intended as an exploratory investigation of its structural and criterion validity (John & Benet-Martinez, 2000). A further aim was to describe its psychometric properties.

As test anxiety is assumed to be a multifaceted construct (Zeidner, 1998), the first step was to analyze the factor structure of the WAESTA scale. As the WAESTA items were theoretically designed to tap either the worry, avoidance or emotionality component of the statistics anxiety construct. As the WAESTA items were theoretically designated to represent each the worry, avoidance, and emotionality component of the statistics anxiety construct, a clear three factor solution appeared to be expectable, at best. However, relevant research findings in the test anxiety field had demonstrated these components being substantially correlated (Deffenbacher, 1980; Cassady & Johnson, 2002; Chin, Williams, Taylor, & Harvey, 2017; Gaye-Valentine & Credé, 2013; Hodapp & Benson, 1997; Hong & Karstensson, 2002; Sarason, 1984; Sparfeldt, Schilling, Rost, Stelzl, & Peipert, 2005; von der Embse, Jester, Roy, & Post, 2018). Furthermore, in certain research contexts dealing with school students’ domain- or subject-specific test anxieties, all worry, avoidance, and emotionally items repeatedly loaded on one common anxiety factor (Faber, 1993, 1995a, 2012b, 2015). Therefore, an accurate prediction of the scale’s ultimate factor structure seemed difficult. Rather the present study should explore the scale’s underlying structure in a most tentative way – and should, thus, take into account three alternatives: a three factor solution separating the worry, avoidance, and emotionality components, a two factor solution with the worry and avoidance items loading on a first factor and the emotionality items loading on a second factor, and a one factor solution subsuming all items. Presuming the avoidance component to represent a specific worry element, the two factor solution definitely revealed a reasonable perspective, in particular. With the reservation of this initially performed analysis, the final version of the WAESTA scale should be determined and its psychometric properties examined.

The second step was to test the scale’s external validity, in particular its relationships with selected convergent and discriminant criterion variables. For this purpose, the validation framework included variables representing the students’ cognitive-motivational orientations and the statistically relevant background.

As relevant cognitive-motivational variables the students’ competence and control beliefs were considered. Both types of beliefs are well proven to regulate the students’ engagement and learning approach in the long term. In particular, they essentially affect the students’ anxiety experience. As unfavorable competence and control beliefs usually come along with increased expectancies of failure, they will provoke a strong sense of personal threat and, thus, lead to an individually heightened level of anxiety. For the purpose of scale validation, as relevant competence beliefs their domain-specific self-concepts, as relevant control beliefs their implicit competence theories were assessed.

There is sound evidence that domain-specific academic competence beliefs and self-concepts substantially predict the individually existing magnitude of test anxiety (Ahmed, Minnaert, Kuyper, & van der Werf, 2012; Faber, 2012a; Goetz, Pekrun, Hall, & Haag, 2006).
Correspondingly, in the statistics domain the crucial role of self-concepts had been well established. In most cases, high-anxious students reported a lowered self-concept of own mathematics or statistical competencies (Bandalos, Yates, & Thorndike-Christ, 1995; Benson, 1989; González, Rodríguez, Failde, & Carrera, 2016; Macher, Paechter, Papousek, & Ruggeri, 2012; Onwuegbuzie & Wilson, 2003; Williams, 2014; Zeidner, 1991). Therefore, it should be assumed the WAESTA scale scores to correlate negatively and substantially with the students’ mathematics self-concept. As well, to sufficiently clarify the domain-specificity of the WAESTA scale, its relation to the students’ verbal self-concept should be concurrently analyzed. According to the multidimensional feature of academic self-beliefs (Green, Martin, & Marsh, 2007; Marsh & O’Mara, 2008), research in the field had consistently demonstrated the students’ mathematics anxiety being substantially related to their performance and motivation in the mathematics but not in the verbal domain (Arens, Becker, & Möller, 2017; Faber, 1995b, 2015; Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007; Gogol, Brunner, Preckel, Goetz, & Martin, 2016; Hembree, 1990; Hill, Mammarella, Devine, Caviola, Passolunghi, & Szücs, 2016; Marsh, 1988; Marsh & Yeung, 1996; Sparfeldt, Schilling, Rost, Stelzl, & Peiper, 2005; Streblow, 2004). Across most diverse educational settings a clear subject-specific pattern of relations between mathematics anxiety and performance or motivation variables had been substantiated – thus evidently indicating the convergent and discriminant validity of mathematics anxiety measures. From this validation perspective, the WAESTA scale should claim preliminary subject-specificity if its correlation with the verbal self-concept variable would turn out to be distinctly weaker than with the mathematics self-concept variable.

Besides, the students’ control beliefs largely manifest as causal attributions to explain individually experienced outcomes (Skinner, 1996) which, in turn, affect their expectancies of future attainment. Students who attribute their failure to stable, internal factors are less likely to strive to improve their competence – for instance, by increasing their learning effort or altering their learning strategies (Weiner, 1985). In a narrow sense, students’ control beliefs may essentially refer to their perceived malleability of own abilities or competencies. Accordingly, the students develop implicit theories or mindsets which may stress an entity view of more or less fixed and unchangeable abilities – or an incremental view of more or less modifiable and changeable abilities (Dweck & Leggett, 1988; Yeager & Dweck, 2012). Though these implicit theories have been demonstrated to correlate with the student’s actual ability level only marginally (Spinath, Spinath, Riemann, & Angleitner, 2003), they significantly affect their motivationally relevant goal orientations, effort beliefs, learning strategies and, eventually, their task performance (Blackwell, Trzesniewski, & Dweck, 2007; Burnette, O’Boyle, VanEpps, Pollack, & Finkel, 2013; Cury, DaFonseca, Zahn, & Elliot, 2008; Jones, Wilkins, Long, & Wang, 2012). These implicit theories principally might not only concern an individual’s cognitive ability but also might emerge in a domain-specific manner and refer to the perceived malleability of certain skills or competencies (Dweck & Molden, 2005) – e.g., in the areas of mathematics and language (Davis, Burnette, Allison, & Stone, 2011; Räty, Kasanen, Kiskinen, Nykky, & Atjonen, 2004; Vogler & Bakken, 2007), foreign language learning (Lou & Noels, 2017), academic writing (Karlen & Compagnoni, 2017), music (Smith, 2005), physical activities (Ommundsen, 2003), biology and science (Chen & Pajares, 2010; Dai & Cromley, 2014). Consequently, they should also play a motivationally crucial role in the students’ learning of statistics. With respect to the statistics domain, an entity view of own
competencies would diminish or even suspend any control perspective. Following the findings of attribution analyses, this loss of control should increase the students’ expectancies of inevitable failure and, thus, should heighten their statistics anxiety (Bandalos, Yates, & Thorndike-Christ, 1995; Carden, Bryant, & Moss, 2004; Covington, 1986; Faber, 2012a; Leppin, Schwarzer, Belz, Jerusalem, & Quast, 1987).

Unfortunately, previous studies in the field had seldom analyzed the role of implicit theories. If at all, they had referred to the students’ general ability beliefs, but not to specific beliefs about statistical competencies (Zonnefeld, 2015) or had only considered the students’ beliefs to master statistical demands through strategy use and effortful behavior (Schultz, Drogosz, White, & Distefano, 1998) – thus, reflecting an incremental view of learning approach. However, these learning control beliefs had been demonstrated to significantly predict course grades. Moreover, in another study the students’ implicit beliefs concerning the malleability of their methodological competencies had been analyzed. Though not directly tapping the issue of statistics, the findings demonstrated a certain motivational type of students showing stable negative perceptions of method learning and, in particular, reporting lowered incremental scores (Lapka, Wagner, Schober, Gradinger, Reimann, & Spiel, 2010).

Accordingly, against the background of overall research findings the WAESTA scale scores should be reasonably assumed to correlate positively and substantially with the students’ entity view of less or not malleable statistical competence.

Furthermore, as another motivational criterion variable, the students’ task values were considered. Conceptually, from an expectancy-value perspective on achievement motivation task values concern the students’ perceived importance or adequacy of a certain activity to fulfill their personal needs and to attain their personal goals (Eccles & Wigfield, 2002). In particular, these task values may manifest as attainment, intrinsic, or utility components and have been demonstrated to regulate the students’ motivational orientations, learning strategies, and academic choices (Wigfield, Hoa, & Klauda, 2009). Task values evidently emerge in a task- or at least domain-specific manner (Gaspard, Häfner, Parrisius, Trautwein, & Nagengast, 2017; Koscovich, Hulleman, Barron, & Getty, 2015; Selkirk, Bouchev, & Eccles, 2011). Thus, they should characteristically affect the students’ learning and performance in the statistics domain as well. Previous research in the field had primarily analyzed the students’ perceived utility or worth of statistical knowledge and competencies as an attitudinal construct as an attitudinal construct (Nolan, Beran, & Hecker, 2012) – focusing on the usefulness of statistics in personal and professional contexts (Cruise, Cash, & Bolton, 1985; Dauphinee, Schau, & Stevens, 1997; Ramirez, Emmioğlu, & Schau, 2010; Schau, Stevens, Dauphinee, & Del Vecchio, 1995). As relevant empirical findings consistently showed, the students’ ratings of the worth of statistics appeared to positively correlate with their statistical achievement as well as with their self-reported learning strategies to a slight extent only (Arumugam, 2014; Emmioğlu & Capa-Aydin, 2012; Hood, Creed, & Neumann, 2012; Kesici, Baloglu, & Deniz, 2011; Nasser, 2004; Slootmaeckers, 2012). In comparison, the relations of utility perceptions with domain-specific measures of academic self-concept were stronger. Students with low competence beliefs valued statistics as less important (Baloglu, 2002; Chiesi & Primi, 2009; Dauphinee, Schau, & Stevens, 1997; Dempster & McCorry, 2009; Stanisavljevic, Trajkovic, Marinkovic, Bukumiric, Cirkovic, & Milic, 2014; Vanhoof, Kuppens, Sotos, Verschaffel, & Onghena, 2011). Similarly, students’ statistics anxiety was also moderately correlated with their utility ratings – indicating those students
suffering from a heightened level of statistics anxiety tendentially perceived statistics to a lesser extent as useful (Baloğlu, 2002; Chew & Dillon, 2014c; Nasser, 2004; Papanastasiou, 2005; Papanastasiou & Zembylas, 2008; Papousek, Ruggeri, Macher, Paechter, Heene, Weiss, Schulter, & Freudenthaler, 2012; Tremblay, Gardner, & Heipel, 2000). In summary, these correlational findings draw a pattern which is completely in line with the social-cognitive processing of task values (Bornholt & Wilson, 2007; Guo, Parker, Marsh, & Morin, 2015). They will operate both as mediating and mediated variables among the students’ actual competencies and self-beliefs (Chiesi & Primi, 2010; Lalonde & Gardner, 1993; Nasser, 2004; Ratanaolarn, 2016; Sesé, Jiménez, Montano, & Palmer, 2015). Accordingly, the WAESTA scale scores should be assumed to correlate inversely and substantially with the students’ perceived value of statistical competence.

As a relevant background variable to explain the students’ statistical self-beliefs and competencies their prior mathematical learning, in particular their latest school grade had been well proven. From the perspective of self-concept development (Marsh & O’Mara, 2008), previous failure experience in mathematics will evidently lead to form low competence beliefs in the statistics domain and, eventually, contribute to strengthening the emergence of domain-specific anxiety responses. In various studies students with poor school grades in mathematics reported a heightened level of statistics anxiety (Beurze, Donders, Zielhuis, de Vegt, & Verbeek, 2013; Birenbaum & Eylath, 1994; Chiesi & Primi, 2010; Galli, Ciancaleoni, Chiesi, & Primi, 2008; Lalonde & Gardner, 1993). Accordingly, the WAESTA scale scores should be assumed to correlate inversely and substantially, but low in magnitude with the students’ mathematics grade they had last earned at school.

Students’ prior experience of statistics should also be considered as a relevant background variable. Following the analysis of Sutarso (1992) who demonstrated significantly lower anxiety scores in the subgroup of students with more statistics courses taken before the exam, Onwuegbuzie (2003) also substantiated a small but significant relation between prior statistics experience and self-concept. Unfortunately, he did not explicitly assess the statistics anxiety variable. Instead he had used the computational self-concept subscale from the STARS instrument (Cruise, Cash, & Bolton, 1985). Although there is theoretical and empirical overlap between the self-concept and anxiety variables they represent distinct constructs and should not be treated as interchangeable. This is particularly important as the STARS subscale does not measure self-concept in a particularly sophisticated way and includes several items that are unrelated to competence beliefs. In another study the students’ previous experiences did not significantly predict their grades in a statistics exam (Slootmaeckers, 2012). These findings seemingly suggest the potential role of prior statistics experiences to explain rather the students’ self-beliefs than their actual attainment. The study of Dempster and McCorry (2009) clearly lent support to this perspective. Being more frequently and persistently involved with learning statistics might increase the students’ confidence with typical requirements and methods (Waples, 2016). Thereby, they might reduce their feelings of apprehension and fear of failure. However, other relevant studies did not substantiate any significant impact of prior statistics experiences on the students’ statistics anxiety (Birenbaum & Eylath, 1994; Chew & Dillon, 2012; 2014b; Schutz, Drogosz, White, & DiStefano, 1998; Townsend, Moore, Tuck,
Most noteworthy, in some cases advanced students’ anxiety level did not decrease after completion of a statistics course or program (Chau, 2018; van der Westhuizen, 2014). In order to further clarify their role to predict the individually existing level of statistics anxiety, the scale validation of the WAESTA scale included the students’ prior statistics experiences. According to the exposure or contact hypothesis, the WAESTA scale scores should be assumed to correlate inversely and substantially, but low in magnitude with the students’ prior statistics experience.

4 Method

4.1 Participants and procedure

In order to conceptually and methodologically extend the scope of preliminary analyses (Faber, Drexler, Stappert, & Eichhorn, 2018), in the present study the data of both a construction sample and a validation sample were further analyzed.

The construction sample consisted of 113 graduate students (n = 94 females, n = 19 males) from a German university Master’s course in educational sciences (n = 80) and special education (n = 33). They all were enrolled in a compulsory course on empirical research methods and theory of science. Therefore, the participation rate was sufficiently high at 82 per cent. Seventy-four of the students had already acquired elementary statistical knowledge during their first degree, whereas 39 were required to attend a course in basic descriptive and inferential statistics. Both the subgroup with and without statistical knowledge did not significantly differ with respect to gender (chi-square test, p > .05) and age (Mann-Whitney U-test, p > .05). Also, there was no significant difference of gender ratio within each subgroup (binomial test, p > .05).

The validation sample was thought to scrutinize the findings from the construction sample one year later. It consisted of 87 graduate students from the same Master’s courses: educational sciences (n = 59) and special education (n = 28). The sample was predominantly female (n = 74). As with the construction sample all the students were enrolled on a compulsory course on empirical research methods and theory of science. The participation rate was rather high at 89 per cent. Fifty of the students had acquired statistical knowledge during their first degree whereas 37 had to attend an introductory statistics course. Once again there were no subgroup differences in gender, gender ratio, or age (Stappert, 2017).

In both samples all relevant data concerning the self-belief and background variables under consideration were gathered on the course’s first term. For that purpose, a questionnaire including all items to measure the students’ self-concept, statistics anxiety, implicit theories, task values, and relevant background information was administered. Whereas the background items as well as the self-concept items were presented separately and blockwise, the other self-belief items were all presented in a mixed order (Schriesheim, Koppelman, & Solomon, 1989). Furthermore, to prevent a priming effect of the self-concept items, they were presented at the end of the questionnaire (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).
4.2 Measures

The students’ academic self-concepts in mathematics and language (German) were assessed using nine six-point rating items for each subject. These items referred to the students’ most recent learning experiences at school and addressed their competence beliefs with regard to meeting subject-specific demands – also in comparison with former classmates. The wording of the items was strictly parallel. In the majority, the items originate from well proven instruments (Faber, 2012a; Möller, Streblow, Pohlmann, & Köller, 2006; Rost, Sparfeldt, & Schilling, 2007). For the purpose of this study they were adapted, phrased retrospectively, and presented in an economical grid style format. Sample item: “I tried hard in mathematics/German, but I did not perform very well.” Exploratory factor analysis (with varimax rotation) revealed a two-factor solution allowing for a clear distinction between the subject-specific self-concept facets. Hence, it was possible to build two scales for measuring the subject-specific academic self-concepts. In both samples, their reliability was most appropriate for both the mathematics and the language self-concept scale (Table 2). According to the multifaceted feature of the construct, the self-concept variables appeared to lowly correlate in the construction sample (r = .14, p > .05) and in the validation sample (r = -.05, p > .05). High scale scores indicated the students’ competence beliefs being positive. For the construction sample, an additionally conducted confirmatory factor analysis using three item parcels for each subject revealed an appropriate fit of the measurement model ($\chi^2$/df = 1.143, TLI = 0.989, CFI = 0.997, RMSEA = 0.036). Here again, the latent mathematics and language variable correlated to a low extent (r = -.25, p < .01). Remarkably, the relation between both self-concept measures turned out to be negatively signed. However, from the perspective of self-concept development, this particular finding might be explainable as it indicates a dimensional comparison effect in the students’ academic self-perceptions across non-matching domains or subjects (Arens, Becker, & Möller, 2017; Marsh, Kuyper, Seaton, Parker, Morin, Möller, & Abduljabbar, 2014).

To assess students’ implicit theory of statistical competencies, a short scale with five four-point rating items was administered in the construction sample. As current instruments in the field only allowed for measuring the students’ implicit intelligence theory (İlhan & Cetin, 2013; Kookén, Welsh, McCoach, Johnston-Wilder, & Lee, 2016), a new scale was created. Following the recommendations of Hong, Chin, Dweck, Lin and Wan (1999), all items tapped an entity view of personal statistical competence. Unfortunately, due to their insufficient item-test correlation ($r_{it} < .22$) two items had been deleted. With an average item intercorrelation of Fisher’s $z' = .44$ and an average item-test correlation of Fisher’s $z' = .52$ the final scale’s reliability appeared to be just acceptable (John & Benet-Martínez, 2000). Sample item: “To work with statistics, you need a talent that I simply do not have.” High scale scores indicated the students to perceive their statistical ability being fixed, hence as less malleable in nature. In the validation sample, a slightly revised scale with four four-point Likert items was used (Stappert, 2017). In view of sample size and item number its reliability appeared to be sufficient (Table 2).

The utility value students attributed to statistical competence was measured by means of a short scale. In the case of the construction sample it consisted of five four-point rating items dealing with the perceived utility of statistics for the students’ current studies and intended
career. To reduce acquiescence or social desirability bias in the students’ responses, scale items were both positively and negatively worded. Sample item: “Statistics will not play an important role in my future professional life”. With an average item intercorrelation of Fisher’s $z' = .40$ and an average item-test correlation of Fisher’s $z' = .49$ the scale’s reliability was just acceptable (Table 2). High scale scores indicated the students to consider statistics being less important. In the validation sample, an extended version of the scale was used (Eichhorn, 2018). It consisted of eight four-point Likert items. Its reliability was once more just acceptable (Table 2). Here again, high sum scores indicated the students to perceive statistics as being less useful for their current studies and later professional development.

Table 2. Descriptive statistics and reliabilities of the scales for measuring self-belief and background validation variables

<table>
<thead>
<tr>
<th>Items</th>
<th>AM</th>
<th>SD</th>
<th>zS</th>
<th>zK</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Latest School Grade in Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction Sample</td>
<td>3.06</td>
<td>0.91</td>
<td>0.76</td>
<td>-0.15</td>
<td></td>
</tr>
<tr>
<td>Validation Sample</td>
<td>3.11</td>
<td>1.36</td>
<td>-0.50</td>
<td>-2.28*</td>
<td></td>
</tr>
<tr>
<td><strong>Prior Statistics Experiences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction Sample</td>
<td>1.15</td>
<td>0.91</td>
<td>1.46</td>
<td>-1.62</td>
<td></td>
</tr>
<tr>
<td>Validation Sample</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Academic Self-Concept in Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction Sample</td>
<td>29.81</td>
<td>9.81</td>
<td>1.37</td>
<td>-2.04</td>
<td>.93</td>
</tr>
<tr>
<td>Validation Sample</td>
<td>31.01</td>
<td>11.43</td>
<td>-0.33</td>
<td>-1.84</td>
<td>.94</td>
</tr>
<tr>
<td><strong>Academic Self-Concept in Language</strong> (German)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction Sample</td>
<td>45.52</td>
<td>7.32</td>
<td>-3.61***</td>
<td>-0.01</td>
<td>.92</td>
</tr>
<tr>
<td>Validation Sample</td>
<td>41.45</td>
<td>8.53</td>
<td>-2.12*</td>
<td>-0.16</td>
<td>.90</td>
</tr>
<tr>
<td><strong>Implicit Entity Beliefs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction Sample</td>
<td>5.43</td>
<td>2.02</td>
<td>3.59***</td>
<td>0.13</td>
<td>.70</td>
</tr>
<tr>
<td>Validation Sample</td>
<td>7.87</td>
<td>2.61</td>
<td>2.15*</td>
<td>-0.32</td>
<td>.81</td>
</tr>
<tr>
<td><strong>Negative Utility Value of Statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction Sample</td>
<td>11.06</td>
<td>2.89</td>
<td>0.41</td>
<td>-1.18</td>
<td>.72</td>
</tr>
<tr>
<td>Validation Sample</td>
<td>16.18</td>
<td>3.95</td>
<td>0.55</td>
<td>-0.10</td>
<td>.75</td>
</tr>
</tbody>
</table>

1In the validation sample, a binary response format was used. Therefore, the item’s mode is reported.

Significance: *p < .05, ***p < .001

Note. AM = arithmetic mean, SD = standard deviation, zS = z-standardized skewness, zK = z-standardized kurtosis, α = internal consistency (Cronbach’s coefficient alpha)

Finally, as relevant background variable students’ most recent school grade in mathematics was inquired in both samples. Students were also asked to report the experience of statistics in the course of their first degree. The construction sample used a four-point Likert scale ranging from “no experience” to “a lot of experience”. As this response format turned out to be too ambiguous in the students’ understanding, in the validation sample their previous experience
with statistics was assessed with a dichotomous format – which referred to the successful passing of a statistics exam (response categories “yes” = 2 and “no” = 1).

4.3 Data analyses

According to the conceptual difficulty to unequivocally predict the WAESTA scale’s ultimate factorial feature, a series of principal component analyses (PCA) was conducted in order to explore and clarify the latent structure underlying the relations among all worry, avoidance, and emotionality items (Kline, 2013). Following the methodological recommendations of Pritch and Stevens (2016), only factor loadings equaling or exceeding an absolute value of $|0.4|$ and, thus, explaining around 16% of the item variance, should be accepted and interpreted – and consequently used for the purpose of final scale formation.

That way, principal component analyses were calculated for the construction sample. The Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett’s test of sphericity demonstrated the inter-item correlations being appropriately strong ($KMO = .878$, $BTS \, p < .001$). In spite of the small number of participants in the validation sample, a principal component analysis of the scale items was also conducted. As the items’ communalities ranged from $h = .412$ to $h = .660$ (Costello & Ossbourne, 2005) and the Kaiser-Meyer-Olkins measure revealed an appropriately high score ($KMO = .901$), this procedure appeared to be most reasonable (de Winter, Dodou, & Wieringa, 2009; MacCallum, Widaman, Zhang, & Hong, 1999).

In order to further clarify possible differences in the statistics anxiety variable and to disclose potential nonlinear relations, a series of two-way analyses of variance (ANOVA) with the students’ mathematics self-concept, implicit entity beliefs, and negative utility values as factor variables was additionally run. In each case, the independent variables were factorized using median split, thus, leading to the identification of lowered and heightened level subgroups. In particular, separate $2 \times 2$ analyses and a comprehensive $2 \times 2 \times 2$ analysis for the construction sample were conducted.

Both samples had missing data (5.7% and 7.3%). As they did not produce any systematic pattern in the construction (MCAR test $p = .182$) and in the validation sample (MCAR test $p = .178$), they were treated as “missing completely at random” (Little, 1988). The missing values were estimated by means of the two-step iterative expectation-maximization (EM) algorithm. Though this method appears to be at risk to underestimate the standard errors of imputed data, in view of the small number of missings, its use was considered justified (Graham, 2012).

5 Results

5.1 Scale formation

For determining the final version of the WAESTA scale in the construction sample, first of all, descriptive item statistics were calculated. To test significant deviations from normal distribution their z-standardized skewness and kurtosis scores were utilized (Field, 2009). The avoidance item 04 as well as the worry item 11 showed a significant negative skew indicating most students to agree with the statements – in detail they would preferably give a presentation without any statistical content and during a presentation they would strongly hope not being
asked statistical questions. Furthermore, as the analysis revealed a significant negative kurtosis score for the items 03, 06, 14, and 15, their distribution appeared to be platykurtic. Accordingly, the students’ relevant item responses denoted a heightened variance or difference among them (Table 3).

Table 3. Descriptive statistics, factor loadings and corrected item-test correlations of WAESTA items (WR = worry, AV = avoidance, EM = emotionality): Results from the construction sample (Faber, Drexler, Stappert, & Eichhorn, 2018)

<table>
<thead>
<tr>
<th>Item</th>
<th>AM</th>
<th>SD</th>
<th>zS</th>
<th>zK</th>
<th>a</th>
<th>rIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 WR</td>
<td>2.37</td>
<td>0.82</td>
<td>1.49</td>
<td>-0.67</td>
<td>.486</td>
<td>.436</td>
</tr>
<tr>
<td>02 EM</td>
<td>2.71</td>
<td>0.94</td>
<td>-1.30</td>
<td>-1.69</td>
<td>.688</td>
<td>.645</td>
</tr>
<tr>
<td>03 AV</td>
<td>2.74</td>
<td>1.08</td>
<td>-1.33</td>
<td>-2.62**</td>
<td>.581</td>
<td>.515</td>
</tr>
<tr>
<td>04 AV</td>
<td>2.88</td>
<td>0.97</td>
<td>-2.12*</td>
<td>1.65</td>
<td>.719</td>
<td>.664</td>
</tr>
<tr>
<td>05 WR</td>
<td>2.51</td>
<td>0.91</td>
<td>-0.50</td>
<td>-1.68</td>
<td>.725</td>
<td>.663</td>
</tr>
<tr>
<td>06 AV</td>
<td>2.50</td>
<td>0.97</td>
<td>-0.21</td>
<td>-2.08*</td>
<td>.758</td>
<td>.697</td>
</tr>
<tr>
<td>07 EM</td>
<td>2.17</td>
<td>0.96</td>
<td>1.71</td>
<td>-1.78</td>
<td>.697</td>
<td>.620</td>
</tr>
<tr>
<td>08 WR</td>
<td>2.24</td>
<td>0.84</td>
<td>1.49</td>
<td>-0.83</td>
<td>.718</td>
<td>.670</td>
</tr>
<tr>
<td>09 EM</td>
<td>2.20</td>
<td>0.87</td>
<td>1.59</td>
<td>-1.04</td>
<td>.668</td>
<td>.583</td>
</tr>
<tr>
<td>10 WR</td>
<td>2.69</td>
<td>0.93</td>
<td>-1.63</td>
<td>-1.43</td>
<td>.647</td>
<td>.694</td>
</tr>
<tr>
<td>11 WR</td>
<td>2.91</td>
<td>0.97</td>
<td>-2.09*</td>
<td>-1.75</td>
<td>.740</td>
<td>.687</td>
</tr>
<tr>
<td>12 WR</td>
<td>2.26</td>
<td>0.80</td>
<td>1.52</td>
<td>-0.45</td>
<td>.734</td>
<td>.618</td>
</tr>
<tr>
<td>13 EM</td>
<td>2.96</td>
<td>0.93</td>
<td>-1.51</td>
<td>-0.45</td>
<td>.683</td>
<td>.493</td>
</tr>
<tr>
<td>14 WR</td>
<td>2.69</td>
<td>1.07</td>
<td>-0.67</td>
<td>-2.18*</td>
<td>.551</td>
<td>.474</td>
</tr>
<tr>
<td>15 EM</td>
<td>2.87</td>
<td>0.84</td>
<td>-1.30</td>
<td>-2.77**</td>
<td>.532</td>
<td>.637</td>
</tr>
<tr>
<td>16 WR</td>
<td>2.75</td>
<td>0.84</td>
<td>-1.45</td>
<td>-0.83</td>
<td>.689</td>
<td>.442</td>
</tr>
<tr>
<td>17 AV</td>
<td>2.07</td>
<td>0.87</td>
<td>1.55</td>
<td>-1.52</td>
<td>.488</td>
<td>.442</td>
</tr>
<tr>
<td>Total Sum Score (Items 01-17)</td>
<td>42.66</td>
<td>10.20</td>
<td>-0.49</td>
<td>-1.34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance: *p < .05, **p < .01

Note. AM = arithmetic mean, SD = standard deviation, zS = z-standardized skewness, zK = z-standardized kurtosis, a = factor loading, rIT = corrected item-test correlation

For the data of the construction sample, principal component analyses did not statistically separate the three anxiety components. Due to the eigenvalues (e1 = 7.385, e2 = 1.219, e3 = 1.132), neither a varimax nor an oblique rotation procedure yielded any loading pattern to separate the worry, avoidance, and emotionality items in a conceptually proper way. Rather all analyses led to a unidimensional structure (Table 3). This solution revealed sufficiently high factor loadings and explained 43.59 per cent of extracted variance. Nonetheless, for further clarification, three provisional subscales representing the students’
worry, avoidance, and emotionality cognitions were formed, and their interrelations as well as their relations with criterion variables were examined. In line with the PCA result, the subscales strongly correlated. Furthermore, using Steiger’s Z formula (Hoerger, 2013; Steiger, 1980) to compare dependent correlation coefficients, no difference between coefficients reached statistical significance (p > .05). Accordingly, the worry, avoidance, and emotionality subscales appeared to be equally associated with the achievement and motivation variables under consideration (Table 4).

Table 4. Correlations between provisory WAESTA subscales and with criterion variables (in the construction sample)

<table>
<thead>
<tr>
<th></th>
<th>AV</th>
<th>EM</th>
<th>SGM</th>
<th>MSC</th>
<th>VSC</th>
<th>IEB</th>
<th>NUV</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAESTA-Worry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAESTA-Avoidance (AV)</td>
<td>.72***</td>
<td>.80***</td>
<td>-.23*</td>
<td>-.38***</td>
<td>.03</td>
<td>.58***</td>
<td>.41***</td>
</tr>
<tr>
<td>WAESTA-Emotionality (EM)</td>
<td>.69***</td>
<td>-.28**</td>
<td>-.43***</td>
<td>.05</td>
<td>.56***</td>
<td>.57***</td>
<td></td>
</tr>
</tbody>
</table>

Significance: *p < .05, **p < .01, ***p < .001  
Note. SGM = Most Recent School Grade in Mathematics, MSC = Mathematics Self-Concept, VSC = Verbal Self-Concept, IEB = Implicit Entity Beliefs, NUV = Negative Utility Value

Consequently all WAESTA items were used to build the scale’s final version. For its total score, neither the z-standardized scores of skewness and kurtosis nor the Shapiro Wilk W-test (W = .988, df = 113, p = .443) evinced any significant deviation from the normal distribution assumption. High total scores indicated the students’ to report stronger worry, avoidance, and emotionality cognitions. The scale’s reliability was estimated in various ways and turned out to be adequate: Its internal consistency (Cronbach’s coefficient alpha) amounted to α = .92, its split-half reliability (odd-even method using Spearman-Brown correction) to r12 = .89, and its standard error (based on coefficient alpha) was se = 2.67.

In the validation sample, the PCA results revealed eigenvalues (e1 = 8.422, e2 = 1.216, e3 = 1.006) which indicated a solution with one common factor. All factor loadings appeared to be considerably high (ranging from αmin = .612 to αmax = .812). Accordingly, the unidimensional scale feature was fully replicated and explained 52.64 per cent of extracted variance. For the total sum score, z-standardized skewness (zS = -0.096) and kurtosis values (zK = -0.173) did not indicate any significant deviation from the normal distribution assumption. Here again, the scale’s reliability was estimated in various ways and turned out to be adequate: its internal consistency (Cronbach’s coefficient alpha) amounted to α = .94, its split-half reliability (odd-even method using Spearman-Brown correction) to r12 = .91, and its standard error (based on coefficient alpha) was se = 2.45.
5.2 Validation results

As an initial approach to analyze the external validity of the WAESTA scale, its zero-order correlations with the criterion variables under consideration were first analyzed. Due to the non-normal distribution of criterion variables, Pearson and Spearman correlation coefficients were precautionally calculated. However, as both types of coefficients revealed quite similar information, the construct relations were preferably described in terms of Pearson’s r. As the results consistently demonstrated (Table 5), the WAESTA scores were closely and significantly associated with the students’ mathematics self-concept but not with their language self-concept. This particular finding might be considered to provide preliminary evidence that the WAESTA scale measures rather a domain-specific than a general facet of the students’ test anxiety experience.

Table 5. Relations of WAESTA sum scores with background and self-belief variables (zero-order correlations): Results from the construction and the validation sample

<table>
<thead>
<tr>
<th>Most Recent School Grade Mathematics</th>
<th>Prior Experiences Statistics</th>
<th>Academic Self-Concept Mathematics</th>
<th>Academic Self-Concept Language</th>
<th>Implicit Entity Beliefs</th>
<th>Negative Utility Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.29***</td>
<td>-.23*</td>
<td>-.43***</td>
<td>.09</td>
<td>.62***</td>
<td>.49***</td>
</tr>
<tr>
<td><strong>Validation Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.22*</td>
<td>-.18*</td>
<td>-.38***</td>
<td>-.08</td>
<td>.80***</td>
<td>.32***</td>
</tr>
</tbody>
</table>

Significance: *p < .05, **p < .01, ***p < .001

Furthermore, the scale scores were most strongly correlated with the students’ entity views of own statistical competence. A heightened level of statistics anxiety came along with a deep understanding of own statistical competencies being less or even not malleable in nature. With the negative instrumental value of statistics the WAESTA sum score correlated moderately positive. Students reporting a higher level of statistics anxiety tendentially perceived statistical competencies as less important. Finally, the relation between the WAESTA score and the most recent school grade in mathematics appeared to be positive and significant, though low in magnitude. Hence, students with a heightened level of statistics anxiety had been less successful in the mastery of mathematical demands at school. Similarly, students with prior experiences with statistics displayed a lower level of statistics anxiety. Overall, the analysis in both the construction and the validation sample revealed a most comparable pattern of criterion-related correlations.

To get most differentiated validation results, a series of regression analyses with the WAESTA scale as dependent variable were computed for both samples. As this procedure allowed for controlling the covariations among all predictor variables with respect to their empirical overlap and multicollinearity (Field, 2009), it should help to unravel the complexity of construct relations. In particular, a sequence of regression models including an advancing number of predictor variables was consecutively tested (Table 6). In both samples the standardized residuals of WAESTA sum scores did not violate the normal distribution assumption. In each case, the Shapiro Wilk W-test demonstrated the standardized residuals being normally
distributed (construction sample: $W = 986, \text{df} = 113, p = .308$, validation sample: $W = 979, \text{df} = 87, p = .182$).

*Table 6.* Multiple regression of WAESTA sum scores on background and self-belief variables (standardized beta weights and squared multiple regression coefficients): Results from the construction and the validation sample

<table>
<thead>
<tr>
<th>Model</th>
<th>Most Recent School Grade</th>
<th>Prior Experiences Statistics</th>
<th>Academic Self-Concept Mathematics</th>
<th>Implicit Entity Beliefs</th>
<th>Negative Utility Value</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.177</td>
</tr>
<tr>
<td>Val. S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.236</td>
</tr>
<tr>
<td>Val. S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.478</td>
</tr>
<tr>
<td>Val. S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.559</td>
</tr>
<tr>
<td>Val. S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance: *$p < .05$, **$p < .01$, ***$p < .001$

*Note.* $R^2$ = adjusted multiple regression coefficient squared

The results for regression models A and B clearly demonstrated the mathematics self-concept to explain the most part of anxiety variance. However, adding the entity beliefs to the regression equation in model C and D, the predictive power of the students’ mathematics self-concept was reduced to a minimal and insignificant extent. Instead, the students’ entity beliefs largely contributed to the WAESTA sum score. As the mathematics self-concept and the entity belief variable in both samples were substantially correlated ($r = -.51$ in the construction sample and $r = -.45$ in the validation sample), but the entity belief variable in both samples was more strongly related to the anxiety variable ($r = .63$ in the construction sample and $r = .80$ in the validation sample) – the massive decline in the self-concepts’ beta weight must be seen as a result of multicollinearity (Marsh, Dowson, Pietsch, & Walker, 2004). This predictive pattern occurred in both samples.

Moreover, only in the construction sample the students’ negative value of statistics substantially and independently explained additional variance in the WAESTA sum score. The difference between samples might be due to the fact, that the methods for assessing the value variable were not comparably formatted. Similarly, it was only in the construction sample that that prior experience of statistics contributed incrementally to variance in WAESTA score.
Students with more experience reported statistics anxiety to a lesser degree. Once again the difference between samples should be traced back to the scale format. In the validation sample, the dichotomous experience score was treated as a dummy variable which might have reduced empirical variance and, thus, lowered correlation coefficients. Apart from this, all regression analyses demonstrated the students’ statistics anxiety to be essentially and most closely predicted by their control beliefs – as reflecting their perceived malleability of individual competencies in the statistics domain.

Finally, for further clarification of the WAESTA scale’s validity, their mean differences between both the educational science and the special education students were analyzed. As the comparison groups were small and unequal in size, the nonparametric Mann-Whitney U-test for independent samples was used (Zimmerman, 1987). In the construction sample a significantly higher level of statistics anxiety in the special education subgroup were found (Z = -2.314, p = .021, effect size r = 0.22). In the validation sample, the level of statistics anxiety did not significantly differ between educational science and special education students (Z = -0.374, p > .05). However, with the small size of the validation sample in mind, this finding should be considered cautiously.

5.3 Additional multivariate analyses

In order to scrutinize the role of the WAESTA scores within the multivariate context of relevant self-belief variables, their mean differences depending on the students’ domain-specific self-concept, implicit entity, and negative value responses were analyzed. As the result of the first two-way analysis of variance demonstrated, mean differences in the students’ statistics anxiety appeared to be significantly explained by a main effect of both the entity belief and the self-concept variable (Table 7). Accordingly, those students reporting a lowered sense of malleability of their statistical competencies and having a poorer mathematical self-concept apparently displayed a heightened level of worry, avoidance, and emotionality cognitions (Figure 1). The interaction effect of the self-concept and implicit entity beliefs variable was statistically not significant.

A quite similar effect pattern was found with regard to the negative value variable. Mean differences in the WAESTA responses were significantly explained by a main effect of both the negative value and the self-concept variable (Table 7). Those students who perceived the availability of statistical competencies as less useful and, besides, reported a poorer mathematical self-concept, showed a heightened level of statistics anxiety (Figure 1). Thus, in both analyses the students’ statistics anxiety appeared to be substantially predicted by distinct effects of the domain-specific self-beliefs under consideration – whereas the effect of the mathematics self-concept was less powerful in each case. Here again, the interaction effect of both independent variables was statistically not significant.

Bringing together both perspectives by concurrently analyzing the role of all self-belief variables, the result of a three-way analysis of variance revealed significant main effects of the entity beliefs and the negative value but not anymore for the self-concept variable (Table 6). Rather, the students with strong entity beliefs and negative utility values reported a considerably heightened level of statistics anxiety (Figure 2).
Figure 1. Mean statistics anxiety scores depending on the students’ implicit entity theory, negative utility value, and mathematics self-concept: separate two-way analyses of variance from the construction sample.

Figure 2. Mean statistics anxiety scores depending on the students’ implicit entity theory, negative utility value, and mathematics self-concept: comprehensive three-way analysis of variance from the construction sample.

Though this effect pattern obviously appeared somewhat more pronounced for the subgroup with a poorer mathematical self-concept, this difference did not reach statistical significance. Therefore, the additive impact of lowered expectations to improve one’s own statistical skills and negative perceptions of the utility of statistical skills allowed for best predicting the interindividually existing level of statistics anxiety. All interaction effects were statistically not significant. This finding is completely in line with the regression results (Table 5). However, it may yield a more descriptive look on the relations among constructs. That is, within the cognitive-motivational interplay of all self-belief variables, the explanatory role of the
mathematical self-concept was largely mediated by the students’ entity beliefs and utility values.

Table 7. Differences in students’ statistics anxiety depending on their mathematics self-concept, implicit entity beliefs, and negative instrumental value. Results from the construction sample.

<table>
<thead>
<tr>
<th>WAESTA Sum Scores (Statistics Anxiety)</th>
<th>Factor Variables</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics Self-Concept (MSC)</td>
<td>1,112</td>
<td>5.749</td>
<td>.018</td>
<td>.050</td>
<td></td>
</tr>
<tr>
<td>Implicit Entity Beliefs (IEB)</td>
<td>1,112</td>
<td>48.671</td>
<td>.000</td>
<td>.309</td>
<td></td>
</tr>
<tr>
<td>MSC X IEB</td>
<td>1,112</td>
<td>0.875</td>
<td>.352</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>Mathematics Self-Concept (MSC)</td>
<td>1,112</td>
<td>7.100</td>
<td>.009</td>
<td>.061</td>
<td></td>
</tr>
<tr>
<td>Negative Utility Value (NUV)</td>
<td>1,112</td>
<td>24.349</td>
<td>.000</td>
<td>.183</td>
<td></td>
</tr>
<tr>
<td>MSC X NUV</td>
<td>1,112</td>
<td>0.260</td>
<td>.611</td>
<td>.002</td>
<td></td>
</tr>
</tbody>
</table>

Note. df = degree of freedom, F = F-value, p = probability, η² = partial eta squared (effect size)

6 Conclusion

The present study should examine the internal and external validity as well as the psychometric properties of the newly developed WAESTA scale for measuring educational science students’ worry, avoidance, and emotionality cognitions in the domain of statistics learning. Conceptually, this measurement approach should integrate both the strengths of a more situation- and a more response-focused research line in the field. Data were gathered in a construction sample as well, one year later, in a validation sample of education science majors. As a substantive result, this scale was demonstrated to represent the construct in a unidimensional manner. Using principal component factor analysis, all scale items were empirically supported. Accordingly, the final scale version included all items as initially administered in both samples. Its internal consistency was most sufficient. Furthermore, its relations with various self-belief and background variables most widely turned out as theoretically predicted. Specifically, total WAESTA score was more strongly correlated with the students’ mathematics than with their language self-concept – and, thus, the scale should claim domain-specific validity. These findings correspondingly held for both the construction and the validation sample. For the time being, the WAESTA scale can be considered internally and externally valid as
well as having adequate psychometric properties. Nevertheless, some results definitely deserve further attention.

In particular, the scale’s underlying structure consistently appeared to be unidimensional. This finding indicates the strong empirical overlap among the worry, avoidance, and emotionality responses – and, thus, the cognitive-motivational interplay of anxiety components. Similarly, close relations had been already found elsewhere (Deffenbacher, 1980; Hodapp & Benson, 1997; Cassady & Johnson, 2002; Chin, Williams, Taylor, & Harvey, 2017; Hong & Karstensson, 2002; Sarason, 1984; Sparfeldt, Schilling, Rost, Stelzl, & Peipert, 2005), especially with respect to domain- or task-specific facets of test anxiety (Faber, 1993, 1995a, 2012b, 2015). Nonetheless, in most cases worry and emotionality cognitions were demonstrated to differently correlate with students’ performance and motivation scores (Faber, 2000, 2002; Seipp, 1991; von der Embse, Jester, Roy, & Post, 2018). By no means, this result does challenge the need for a separate assessment of worry, avoidance, and emotionality cognitions. Rather this approach should ensure to obtain a most differentiated measuring of statistics anxiety and, thereby, should contribute to reducing the interpretation ambiguity of item responses. In that regard, it should certainly increase the scale’s cognitive-motivational representativity and content validity. Likewise, the students’ avoidance cognitions were found to be most closely related to their worry, but only slightly less closely related to their emotionality cognitions. Therefore, according to relevant findings (Galassi, Frierson, & Sharer, 1981; Hagtvet & Benson, 1997), avoidance cognitions must be seen as an important feature within the students’ anxiety experience and, thus, should have contributed to completing and differentiating the measuring of statistics anxiety. Hence, future research should more strongly pay attention to this issue and reconsider its theoretical framework as well as its methodological implications. Apart from this conceptual perspective, the WAESTA scale’s content and format should be further refined. In particular, the proportion of worry, avoidance, and emotionality items should be more balanced – in particular, the number of avoidance items should be extended. Furthermore, given the ambiguous research findings (Lee & Paek, 2014; Lozano, García-Cueto, & Muñiz, 2008; Wakita, Ueshima, & Noguchi, 2012; Weng, 2004), subsequent scale analyses should also compare various scale versions with different numbers of response categories in order to examine their psychometric and validity properties.

With respect to the validation results, both the correlation and regression analyses suggest, at first glance, that students’ implicit entity beliefs are sufficient to explain their statistics anxiety. Without any doubt, the entity beliefs appear to obviously play a crucial role in the prediction of statistics anxiety – as should be expected from the view of social-cognitive theories (Dweck & Leggett, 1988; Schunk & Zimmerman, 2006; Zeidner, 1998). However, in the students’ cognitive-motivational processing, they will operate in a more complex manner. According to relevant theoretical conceptions and empirical findings in the field (Blackwell, Trzesniewski, & Dweck, 2007; Chiesi & Primi, 2010; Emmioğlu, 2011; Onwuegbuzie, 2003; Sesé, Jiménez, Montano, & Palmer, 2015), it should be assumed that implicit beliefs actually mediate the effects of students’ self-concept and their learning background on the dependent anxiety variable (Baron & Kenny, 1986). The massive decline in the self-concept variable’s beta weights, when adding the entity belief variable to the regression equation, apparently
supports this assumption (Table 5). The same applies for the results of additionally run analyses of variance (Table 6). Indeed, within the validation framework this indirect effect cannot be adequately substantiated with correlational or regression analysis, but only with multivariate modeling method (Kline, 2011). Future research should make every effort to apply such modelling techniques in order to clarify the role of entity beliefs in the statistics domain.

Beyond the purpose of scale validation, the empirical findings concerning the students’ entity beliefs might even extend the previous research in two respects: at the level of construct specificity, the measuring of entity beliefs did not refer to the perceived malleability of general cognitive abilities but enquired the perceived malleability of statistical competencies. As this entity belief variable was only very weakly correlated with language self-concept (construction sample $r = .07$; validation sample $r = -.12$) it should be considered domain-specific. Hence, for the domain of statistical learning, this particular finding appears to be in line with the recommendations of the implicit theories approach (Dweck & Molden, 2005). Accordingly, at the level of construct relations, the results allow for refining the nomological scope of the statistics anxiety framework – at least, as it refers to the type and role of self-belief variables (Bandolos, Yates, & Thorndike-Christ, 1995; González, Rodríguez, Falide, & Carrera, 2016; Onwuegbuzie & Wilson, 2003; Zeidner, 1991).

The present study undeniably suffers from some conceptual and empirical limitations and shortcomings. First of all, composition and size of both student samples do not allow for generalizing the empirical findings as should be required. Instead, the findings reported here might claim a sort of local validity – all the more, as their data basis referred to a certain university setting. Their external validity must, therefore, be considered weak. Further analyses should necessarily remedy this problem and examine the WAESTA scale with other student samples from other educational science contexts.

Moreover, the validation framework is still lacking in several respects and should be further completed (Benson, 1998). The number of items for measuring the students’ implicit entity theory and utility value should be extended – not least, in order to enhance their internal consistencies. However, simply increasing the number of items will not necessarily improve reliability and criterion validity to a considerable extent (Gogol, Brunner, Goetz, Martin, Ugen, Keller, Fischbach, & Preckel, 2014). Further validation efforts should carefully explore options to broadening the constructs’ operationalizations in a conceptually and contextually adequate manner – but not only to enhance statistical variance of item responses. Especially the construct of entity beliefs might allow for response variability merely in its context or situation facet. Therefore, as relevant studies in the field correspondingly showed, its potential operationalization space will be conceivably restricted (Blackwell, Trzesniewski, & Dweck, 2007; Hong, Chiu, Dweck, & Wan, 1999; Lou & Noels, 2017). One way to refine the construct should be its extension with the students’ beliefs about perceived effort and learning control (Blackwell, Trzesniewski, & Dweck, 2007; Schutz, Drogosz, White, & Distefano, 1998; Tempelaar, Rienties, Giesbers, & Gijselaers, 2015). Furthermore, this study assessed students’ mathematics self-concept retrospectively. As a proportion of both samples did not have any prior experience of statistics using a measure of statistical self-concept would have been misguided. However, provided that further research would include participants being most comparable in their statistical background, their self-concept in the statistics domain should
be absolutely used to elaborate scale validation (González, Rodríguez, Failde, & Carrera, 2016; Sproesser, Engel, & Kuntze, 2016). Likewise, concurrent measures of the students’ self-efficacy to master certain statistical tasks should help to further differentiate the scale’s criterion validity (Finney & Schraw, 2003; Olani, Harskamp, Hoekstra, & van der Werf, 2010; Perpeczkza, Chandler, & Becerra, 2011). Not least, an appropriate validation of the WAESTA scale will require an analysis of its relations with other instruments for measuring statistics anxiety – for instance, by comparing it with the German adaptation of the STARS questionnaire (Papousek, Ruggeri, Macher, Paechter, Heene, Weiss, Schulter, & Freudenthaler, 2012).

As another crucial point the measuring of the students’ prior statistics experience might be considered. In both samples, this variable had been differently operationalized. Therefore, its relations to the statistics anxiety score cannot be compared across samples. However, the multivariate findings indicate the students’ informally experienced exposure to statistics (as assessed in the construction sample) being stronger related to their anxiety scores than their prior passing of a formal exam (as assessed in the validation sample). Possibly, this particular result should be traced to both the binary item format and the number of students in the validation sample. Moreover, in line with research findings reported elsewhere (Dempster & McCorry, 2009), this particular result might also indicate the everyday encounter with statistical demands to play a more important role for the emergence and maintenance of anxious reaction patterns. Further validation efforts should bear this perspective in mind and always analyze the informal experiences with statistics as cognitive-motivationally decisive background variable – instead of merely focussing on structural information (Hannigan, Hegarty, & McGrath, 2014; Onwuegbuzie, 2003). However, as additional analyses of the data from both samples showed, the relation between students’ prior statistics experiences and their statistics anxiety appeared to be largely superposed by the domain-specific self-concept and value variable (Faber & Drexler, 2018).

Another considerable lack of the present study concerns the missing of a relevant performance measure. As only a certain part of students in both samples had yet to pass an exam in introductory statistics, sufficiently robust data were not available. Further validation studies should analyze the relation between the WAESTA scores and suitable measures of students’ actual statistics performance (Hanna & Dempster, 2012). Especially, this relation should be most instructive – in as much as relevant studies commonly reported low to moderate correlations (Bandalos, Yates, & Thorndike-Christ, 1995; Finney & Schraw, 2003; Macher, Paechter, Papousek, & Ruggeri, 2012; Sesé, Jiménez, Montano, & Palmer, 2015; Tremblay, Gardner, & Heipel, 2000; Vigil-Colet, Lorenzo-Seva, & Condon, 2008; Zeidner, 1991). However, these results do not really indicate a general flaw in the measures’ criterion validity. Rather, they reflect the motivational consequences of statistics anxiety within a strongly restricted setting (Pekrun, 1988). As the successful passing of statistical requirements in the Master’s degree is mandatory, the students’ increasingly experienced worry, avoidance tendencies, and feelings of apprehension should dispose them to strengthen their learning effort in order to avoid an impending failure outcome (Macher, Papousek, Ruggeri, & Paechter, 2015; Martin & Marsh, 2003). Accordingly, for the WAESTA scale, also a moderate relation with the students’ statistical performance should be assumed. Moreover, in order to further elucidate the relation between students’ statistics anxiety and actual achievement, it should be also enlightening to
analyze their domain-specific learning approach (Chiesi, Primi, Bilgin, Lopez, del Carmen Fabrizio, Gozlu, & Tuan, 2016; Markle, 2017).

Finally, as both samples in this study were small and predominantly female, gender was not included in the validation analyses. Relevant findings in the field consistently demonstrated the females to report a higher level of statistics anxiety (Benson, 1989; Hong & Karstensson, 2002; Macher, Paechter, Papousek, & Ruggeri, 2012; Onwuegbuzie & Wilson, 2003). Interestingly, despite the apparently heightened anxiety level of female students, some studies did not substantiate any significant disadvantage in their exam performance (Bradley & Wygant, 1998; Macher, Paechter, Papousek, Ruggeri, Freudenthaler, & Arendasy, 2013). This finding needs further clarification with respect to the underlying motivational and behavioral processes. Hence, female students might have overrated their individually existing anxiety level (Zeidner, 1998) – possibly due to a self-derogatory gender stereotyping effect (Bieg, Goetz, Wolter, & Hall, 2015; Pomerantz, Altermatt, & Saxon, 2002). As well, pursuing a more adaptive coping strategy to avoid feared failure, they might have ramped up their learning approach (Martin & Marsh, 2003). Given a larger sample size with a more adequately balanced gender ratio, this issue should also be examined with respect to the WAESTA scale.

In summary, the present findings yield important information concerning the internal and external validity of the newly developed WAESTA scale. However, they must be seen as preliminary in nature. Therefore, they should represent just a very first step in method development.

7 References


Measuring Education Science Students’ Statistics Anxiety


Eichhorn, J. (2018). Die Erfassung von studentischen Valenzkognitionen zur Nützlichkeit statistischer Kompetenzen. Entwicklung, psychometrische Prüfung und vorlängige Validierung eines entsprechenden Befragungsinstruments [Measuring educational science majors’ utility value of statistical competence. Scale development, psychometric properties, and preliminary validation results]. Leibniz University Hannover, Faculty of Humanities (Institute


1382.


and Zeitlinger.


8 Appendix

WAESTA scale for measuring education science students’ worry (WR), avoidance (AV), and emotionality (EM) cognitions encountering statistical demands (from Faber, Drexler, Stappert, & Eichhorn, 2018) – originally worded in German (Faber, Drexler, & Stappert, 2018).

01 WR  I will hardly be able to meet my degree program's statistics requirements right away.
02 EM  I would be very uncomfortable if I had to work on a statistical problem.
03 AV  If I could, I would rather take two other courses than do one statistics course.
04 AV  When presentation topics are being assigned in the course, I would make sure that I receive a topic that doesn't involve statistics.
05 WR  It would be difficult for me to discuss statistical content adequately in my papers.
06 AV  When preparing presentations, I would rather omit anything that has to do with statistics.
07 EM  I would be quite nervous if I were asked to explain a chart from a research report.
08 WR  I have difficulty understanding statistical content in a lecture.
09 EM  I would have trouble extracting the relevant information from a table of statistical values.
10 WR  If I had to comment on statistical data in a course, I would be worried that I would make a fool of myself.
11 WR  If I had to give a presentation including statistical findings in a course, I would hope that no one had any follow-up questions.
12 WR  I would hardly be able to present a report on statistical research findings adequately.
13 EM  I would feel very tense if I had to apply a statistical formula.
14 WR  Despite careful preparation for a statistics exam, I would worry about not passing it.
15 EM  The thought of having to explain a statistical problem in a course makes me quite nervous.
16 WR  If I took a statistics course, I would be concerned that I would quickly forget everything I had learned.
17 AV  In scientific texts, I would skip over statistical tables and diagrams if possible.