CrossMark

The successful test taker: exploring test-taking behavior profiles through cluster analysis

Tova Stenlund[1] · Per-Erik Lyrén[2] · Hanna Eklöf[2]

Abstract To be successful in a high-stakes testing situation is desirable for any test taker. It has been found that, beside content knowledge, test-taking behavior, such as risk-taking strategies, motivation, and test anxiety, is important for test performance. The purposes of the present study were to identify and group test takers with similar patterns of test-taking behavior and to explore how these groups differ in terms of background characteristics and test performance in a high-stakes achievement test context. A sample of the Swedish Scholastic Assessment Test test takers ($N = 1891$) completed a questionnaire measuring their motivation, test anxiety, and risk-taking behavior during the test, as well as background characteristics. A two-step cluster analysis revealed three clusters of test takers with significantly different test-taking behavior profiles: a *moderate* ($n = 741$), a *calm risk taker* ($n = 637$), and a *test anxious risk averse* ($n = 513$) profile. Group difference analyses showed that the *calm risk taker* profile (i.e., a high degree of risk-taking together with relatively low levels of test anxiety and motivation during the test) was the most successful profile from a test performance perspective, while the *test anxious risk averse* profile (i.e., a low degree of risk-taking together with high levels of test anxiety and motivation) was the least successful. Informing prospective test takers about these insights can potentially lead to more valid interpretations and inferences based on the test scores.

✉ Tova Stenlund
  tova.stenlund@umu.se

  Per-Erik Lyrén
  per-erik.lyren@umu.se

  Hanna Eklöf
  hanna.eklof@umu.se

[1] Department of Psychology, Umeå University, 901 87 Umeå, Sweden

[2] Department of Applied Educational Science, Umeå University, 901 87 Umeå, Sweden

 Springer

Standardized high-stakes achievement tests often present a very structured format, with strict time limits, a component of speededness, a large number of items, often of multiple-choice character, and consequences attached to the test result. Thus, tests like these introduce the test takers to a situation where they might feel more or less comfortable, and to be successful in such test situations, sufficient content knowledge might not be enough. Components such as personal characteristics and appropriate reactions, behaviors, and strategies when taking the test may also be important features. For example, previous studies have quite consistently shown that for a test taker to be successful when taking a test, it is important to be able to reduce anxiety and sustain motivation (Dodeen et al. 2014; Naylor 1997; Sternberg 1998), as well as using effective test-taking strategies, such as willingness to take risks (Bicak 2013; Bond and Harman 1994; Dodeen 2008). Studies in this area often compare groups of test takers and have for example found that high achievers tend to report using more effective test-taking strategies when compared to low achievers (Stenlund et al. 2017; Ellis and Ryan 2003; Hong et al. 2006; Kim and Goetz 1993), that males are more prone to taking risks when answering test items (see, e.g., Baldiga 2014), and that females and low achievers seem to experience higher levels of test anxiety than males and high achievers (Stenlund et al. 2017; Cassady and Johnson 2002; Naylor 1997). Findings like these can help understand and to some extent possibly explain performance differences that are often observed between manifest groups in achievement tests. Still, considering the consequences of successful test-taking behavior in high-stakes test situations, and assuming that test takers adopt different test-taking behavior, exploring differences across groups, in terms of profiles, and identifying patterns that seem associated with sucessful and less successful test-taking, respectively, might add important knowledge to this area. With this study, we therefore aim at identifying subgroups of test takers with similar patterns of the variables risk-taking, motivation, and test anxiety and exploring how they differ in terms of background characteristics and test performance in a high-stakes achievement test context.

Risk-taking

Risk-taking as a general concept is related to a person's probability to engage in risky activities and is defined as "engagement in behaviors that are associated with some probability of undesirable results" (Boyer 2006, p. 291). Risk-taking can be seen as negative as well as positive behavior. Thus, even though avoiding risky situations and activities is an important skill to develop and apply in certain situations, risk-taking is identified as an important strategy when taking a test, in particular speeded multiple-choice tests (Baldiga 2014; Alnabhan 2002). Risk-taking behavior when taking a test is related to test-taking strategies commonly called test wiseness, whereas it addresses the willingness to guess when answering a question where the correct answer is not obvious to the test taker. For example, to use elimination processes or look for clues in the item stem when guessing, or if no clues are to be found to make an educated or intelligent guess. Using these types of guessing strategies might lead to better results regardless of ability level and better results than other students at the same ability level not using these types of strategies (Dodeen 2008). It has been found that the willingness to

guess (i.e., risk-taking) has increased over the years, and that changed test-taking behaviors, such as to use guessing and effective elimination processes to increase the odds of choosing a correct answer, even might be a reasonable explanation to the secular gain in measured IQ over time, that is the so-called Flynn effect (Woodley et al. 2014).

Guessing is, however, not always effective. Just picking an answer option at random is not considered to be an especially successful strategy when taking a test (Ellis and Ryan 2003), but perhaps better than to skip an item. Skipping items instead of guessing has been shown to be an ineffective strategy to use (Baldiga 2014). Test takers who are risk averse (i.e., reluctant to answer under uncertainty) may be more prone to skip items or to spend too much time on difficult questions. Research has also shown that there might be gender differences in risk-taking behavior with females as the more risk averse group (Baldiga 2014; Ben-Shakhar and Sinai 1991; Hirschfeld et al. 1995). For example, Baldiga (2014), who examined gender differences in risk-taking, found that females have a tendency to skip items instead of guessing to a higher extent than males. However, in a recent study, females reported guessing to a higher extent when compared to males (Stenlund et al. 2016), showing the importance of more research in the area. It is also possible that there are differences between different age-groups or groups with different education levels, as research suggest that there might be a maturity effect on test-taking strategies (Geiger 1997).

Test motivation and anxiety

Besides important test-taking variables such as willingness to take risks and make educated guesses, earlier studies have also found that the ability to tackle emotional and motivational factors, such as test anxiety and test motivation, are important for test performance (Cheng et al. 2014; Zeidner 1998). Test anxiety often includes worries and irrelevant thoughts which might affect the possibility to concentrate and the level of understanding and retrieval when undertaking a test (Carter et al. 2008). Consequently, high levels of test anxiety might interfere with optimal test performance (Hembree 1988; Seipp 1991). Low levels of motivation when undertaking a test might also affect test performance (van Barneveld 2007; Wise and DeMars 2005). Motivation and anxiety are thus assumed and have empirically been shown, to have opposite effects on performance, and Wolf and Smith (1995) showed that a high level of motivation coupled with a low level of test anxiety is a desirable combination when it comes to performance. However, as motivation increases, test anxiety also tends to increase, possibly canceling the positive effect of high motivation out (Wolf and Smith 1995; Smith and Smith 2002). The main assumption in high-stakes test contexts is that the level of motivation among test takers is generally high due to the possible positive consequences of a good performance. However, a recent study conducted in a high-stakes test situation showed that high and low achievers differ in reported motivation, with low achievers reporting significantly lower levels of motivation than high achievers (Stenlund et al. 2017). Further, studies have shown that there seems to be gender differences in both reported test anxiety and motivation, where females tend to report higher levels of test anxiety (Stenlund et al. 2017; Cassady and Johnson 2002; Naylor 1997; Zeidner 1998), and males tend to report lower levels of motivation (DeMars et al. 2013).

Test anxiety and test motivation have also been found to be related to general test-taking strategies, including risk-taking (see, e.g., Dodeen et al. 2014; Hong et al. 2006; Peng et al. 2014). The use of effective test-taking strategies has been shown to be positively related to

level of motivation, while test anxiety has been shown to affect the use of test-taking strategies negatively (Dodeen et al. 2014; Peng et al. 2014). Test anxious individuals have been reported to use poorer study and test-taking skills (Culler and Holahan 1980; Huntley et al. 2016). Studies have also shown that test anxiety might be related to whether a test is considered to be difficult or not (Hong et al. 2006), and predicted difficulty as a source of anxiety might be due to either a lack of preparation strategies or as a consequence of low ability (Cassady and Johnson 2002).

In sum, earlier research has found that test-taking behavior in terms of risk-taking and ability to manage emotional and motivational factors are important when taking a test, especially in a high-stakes test situation. However, most studies have investigated these variables separately, and it is therefore not clear whether distinguishable profiles with different combinations of these variables can be observed, and if so, whether different profiles are associated with different levels of performance. To increase the understanding of the successful test taker, the patterns of these variables, whether different profiles have different background characticts and whether they are associated with different levels of performance, need to be examined further.

The purpose of the present study is therefore to classify individuals from the heterogeneous population of the Swedish Scholastic Assessment Test (SweSAT) test takers into homogenus subgroups or profiles, based on their risk-taking behavior, test motivation, and test anxiety. Further, the purpose is to compare these subgroups in relation to demographic variables and test-specific characteristics, such as test performance. More specifically, the research was guided by the following research questions:

– Is it possible to, through cluster analysis, identify distinct clusters of test takers, based on the variables risk-taking behavior, test motivation, and test anxiety?
– If clusters are identified, are there differences between the clusters when it comes to background variables such as gender and number of previous SweSAT tests taken?
– If clusters are identified, are there performance differences between the different clusters, i.e., are different test-taking behavior profiles more or less successful in the SweSAT test-taking situation?

Based on the review of previous research on the variables used for clustering, it seems reasonable to assume that different clusters would have different demographic characteristics and show performance differences, for example that a cluster with many test anxious test takers would consist of more females and perform worse than a cluster with fewer test anxious test takers, while a cluster with many risk-taking individuals and repeaters would perform better. However, as the study is exploratory in nature and clusters not specified in advance, no strict a priori hypotheses were specified.

The SweSAT is a multiple-choice test, including a verbal and a quantitative part. The test is used for selection to higher education in Sweden, which implies that the testing situation in general is viewed as high-stakes for the test takers. Admission regulations stipulate that in a situation where selection has to take place (i.e., there are more applicants than available places), applicants are to be admitted on the basis of either upper-secondary school grades or SweSAT scores, not on a composite of the two. This means that it is not mandatory to take the test if you want to apply to higher education (and on the other hand, there is no limit on the number of times the test can be repeated). Taking the test may increase an applicant's chance of being admitted, and for many applicants, the SweSAT is their only chance of being admitted.

This is true not only for those with poor grades in general, but also for those with very good grades who apply to highly competitive programs and courses. So, the test is generally viewed as high-stakes, but the actual stakes, and therefore the level of motivation and anxiety related to the test-taking experience, may vary between test takers. This notion together with the fact that the test consists only of multiple-choice items makes test motivation, test anxiety, and risk-taking important issues to consider with respect to test taker characteristics and performance in the SweSAT context.

Method

In the present study, cluster analysis is used to place individuals into groups or profiles based on their similarities. In cluster analysis, the main purpose is to group observations by taking distance and similarities into consideration, thus, bring the differences between clusters and the similarities within clusters to the uppermost level. Apart from this, the goal is to make up a model that will determine profiles or types among the subjects or participants.

Participants

A questionnaire was sent out to a random sample (n = 6304) of the individuals that had registered for the SweSAT in the autumn of 2013. Of the 2299 (36.5% of the total sample) that responded to the questionnaire, 155 did not take the SweSAT, and 253 did not answer all questions of the scales forming the clusters. Only data from the respondents who completed the SweSAT and the scales used in the cluster analysis were considered to be relevant for this study (n = 1891). The mean age among the participants was 22.2 years (SD = 6.6), 60% were females, and the mean scores (on a number-correct scale of 0–80) were 38.8 (SD = 12.7) on the quantitative section and 47.1 (SD = 13.6) on the verbal section. The corresponding numbers for the population are 21.6 years (SD = 5.6), 53% females, quantitative score 36.8 (SD = 12.3), and verbal score 43.6 (SD = 13.6). So, the respondents were older, $t(1890)$ = 4.32, $p < .001$ (Cohen's d = 0.11), had a larger proportion of females, z = 6.10, $p < .001$ (Cohen's h = 0.14), and had a higher quantitative score, $t(1888)$ = 7.91, $p < .001$, (d = 0.16) as well as a higher verbal score, $t(1890)$ = 12.62, $p < .001$ (d = 0.26). Thus, the sample differs in a statistically significant way from the study population with respect to age, gender distribution, and performance on the SweSAT. Yet, because the effect sizes are small for all four variables, we would expect the practical significance of the difference to be small and hence the results to be generalizable from the respondents to the population.

Procedure and instrument

The data used in the present study were collected through a post-test, self-report, Web-based questionnaire used in the autumn 2013 administration of the SweSAT. The questionnaire was developed as a means of measuring test takers' perceptions of different aspects of the SweSAT, such as perceived difficulty and relevance, test-taking strategies, motivation, and test anxiety. The questionnaire was open for 4 weeks, from 3 days after the test was administered until 2 days before the test takers could get their online score report. Two reminders were sent out, the first after 6 days and the second after another 6 days. The majority of the participants responded before the first reminder (as the focus of this study is on risk-taking behavior and

other non-cognitive variables that may be important for test performance, the part of the questionnaire evaluating the SweSAT is not relevant for this study and therefore not reported in this text). The data that were used in this particular study are presented in the following.

Scales used to form the clusters: risk-taking, test anxiety, and test-taking motivation

The risk-taking scale included five items asking about different aspects of risk-taking behavior. The items were about how prone the test takers were to guess (e.g., I guessed when I did not know the answer) and whether they favor more secure test-taking strategies, such as spending a lot of time on difficult questions. Test anxiety was measured with seven items, asking primarily for emotional and cognitive aspects of test anxiety such as fear of failing the test (e.g., I was afraid that I would fail the test), worries about the difficulty of the test, and whether the test situation as such made the test taker feel stressed or nervous. Perceived importance and motivation to spend effort on the test were measured with seven items about whether the test taker felt motivated to do his or her best (e.g., I was motivated to do my best at the test), whether a good result was perceived as important, and how much effort the test taker spent on the test. All items in the three scales were rated on a 5-point Likert-type scale ranging from strongly disagree to strongly agree (see Table 1 for descriptive statistics). The anxiety and motivation scales were adapted from previous studies in different Swedish assessment contexts (Eklöf and Nyroos 2013; Knekta and Eklöf 2015), where they have demonstrated acceptable psychometric properties. The risk-taking scale was developed in agreement with earlier research in the area.

In this study, Cronbach's alpha was used to examine the internal consistency reliability of each subscale (Table 1). The analyses showed that the motivation and the test anxiety scales had acceptable values of Cronbach alpha, while the risk-taking scale had a lower alpha value. However, this scale only includes five items, of which some are reversed scored, and therefore, the alpha value can be regarded as tolerable (Field 2013). Further, as the distance measure on which the cluster analysis is based gives the best results if all continuous variables used are independent and have a normal distribution, these aspects were also examined. A descriptive analysis showed that scores from all three scales were normally distributed (see Table 1, skewness did not exceed ±1), and a correlation analysis revealed that the inter-correlations between the three scales were in the low to medium range (see Table 1).

Variables used to describe and compare the clusters

Background characteristics and test specific characteristics were used to describe and compare the clusters. The background variables were gender, age, and education level. The test-specific

Table 1 Descriptive statistics of each scale and intercorrelations ($n = 1891$)

Scale (number of items)	Min-Max	Median	Mean (SD)	Cronbach's alpha	Skewness/kurtosis
Risk-taking (5)	5–25	19.0	18.3 (3.3)	.52	−.420/−.128
Motivation (7)	7–35	21.0	22.2 (5.5)	.79	.039/−.764
Test anxiety (7)	7–35	22.0	20.5 (6.8)	.86	.025/−.493
Intercorrelations	A	B	C		
Risk-taking (A)	–				
Motivation (B)	−.16	–			
Test anxiety (C)	−.42	.35	–		

variables were achievement level, that is, result on the SweSAT, number of tests completed, perceived difficulty of the test, and test preparation (see Table 3 for descriptive statistics for these variables).

Statistical analysis

A two-step approach to cluster analysis was applied, using IBM SPSS Statistics 22™, to cluster the study sample based on the variables risk-taking, test motivation, and test anxiety. The two-step method identifies pre-clusters in the first step and uses hierarchical clustering, treating the pre-clusters as single cases, in the second step. Descriptive statistics (means, *SD*, and standardized means) were used to present the profiles of the clusters and to explore whether the clusters differed in reported risk-taking, motivation, and test anxiety; one-way analyses of variance (ANOVA) were conducted. Further, descriptive statistics, Chi-square test, and one-way ANOVA were used to describe and compare the participants' demographic characteristics and test-specific characteristics in the different groups determined by the cluster analysis. The alpha value was set to .05 for all analyses.

Results

The two-step cluster analysis yielded three clusters based on Schwarz's BIC and the highest Log-likelihood distance measures (ratio of distance measures = 1.738). The smallest cluster has 27.1% of the cases, and the largest has 39.2% (ratio of sizes = 1.44). In clusters 1, 2, and 3, there were 741 (39.2%), 637 (33.7%), and 513 (27.1%) SweSAT test takers, respectively. The three groups were formed based on the similarity of their reported test-taking behavior (i.e., their responses to the items included in the scales of risk-taking, motivation, and test anxiety). In Fig. 1, a description of the profiles of the three clusters and their results in the three variables are presented.
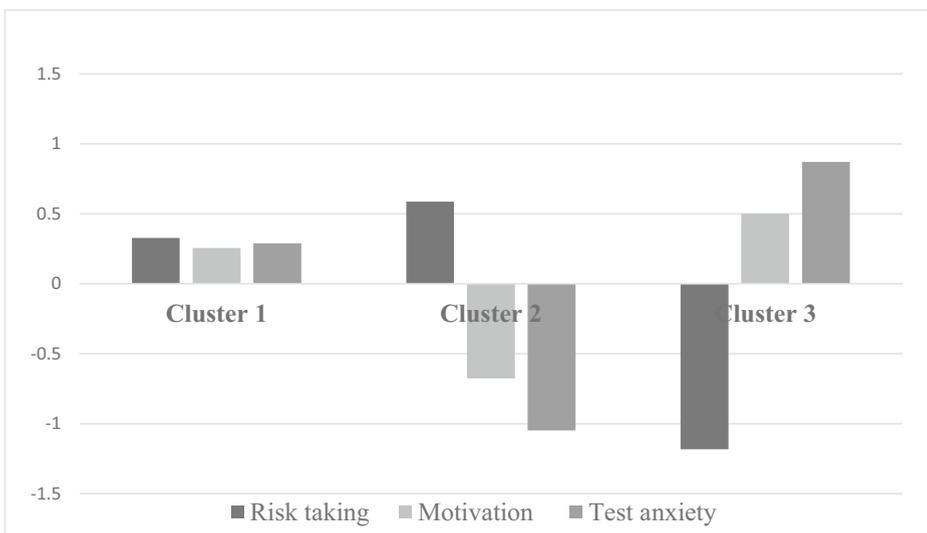

Fig. 1 Profiles of the three clusters (standardized means of risk-taking, motivation, and test anxiety)

Based on their results in the three variables, cluster 1 was labeled *moderate*, cluster 2 was labeled *calm risk taker*, and finally, cluster 3 was labeled *test anxious risk averse*. The labels are based on the relative differences between the clusters rather than on absolute values in the three cluster variables. The clusters differ significantly with regard to reported risk-taking, F (2, 1888) = 1079, $p < .001$, $\eta_p^2 = .53$; motivation, F (2, 1888) = 306, $p < .001$, $\eta_p^2 = .24$; and test anxiety, F (2, 1888) = 1470, $p < .001$, $\eta_p^2 = .60$. Tukey–Kramer post hoc pairwise comparisons showed significant differences between all three clusters in all three variables ($p_s < .001$, see Table 2 for M and SD).

Cluster 1, *moderate*, is the largest of the three clusters, and test takers in this cluster tended to report that they were close to moderately motivated and that they were moderately anxious during the test. Further, this group reported that they used risk-taking strategies to a relatively high degree. Cluster 2, *calm risk taker*, reported low levels of test anxiety and motivation when compared to the other clusters, but reported using risk-taking strategies to a higher extent than the other two clusters. Cluster 3, *test anxious risk averse*, reported the highest levels of test anxiety and motivation and the lowest use of risk-taking strategies when compared to the other clusters. To examine what further characterized the three clusters, the clusters' demographic characteristics (age, gender, and education level) and SweSAT-related characteristics (result on the SweSAT, number of retests, perceived difficulty of the test, and level of preparation before the test) were compared (Table 3). Below, findings for the respective clusters are presented in some more detail.

Cluster 1, moderate

In this cluster, about 67% are females, and it is thereby the cluster with the largest proportion of females. The proportion of females is significantly larger compared to cluster 2 (see Table 3). About 80% are younger than 25 years. Further, in this cluster, about 15% have a higher education (the clusters do not differ in educational level). The mean SweSAT scores for this cluster (total raw score, as well as verbal and quantitative raw scores) are somewhat higher than the scores for cluster 3 (only the verbal score is statistically significantly higher though), and significantly lower than the scores for cluster 2. About 38% of this group are repeaters. Thus, they have completed the SweSAT at least once before this occasion. The proportion of repeaters is significantly smaller when compared to cluster 3. Test takers in this cluster differed significantly from the other two clusters in how difficult they perceived the test (i.e., they perceived the test as more difficult than cluster 2 and less difficult than cluster 3, see Table 3). Finally, the test takers in this cluster perceived themselves as well prepared (no difference between the clusters).

Table 2 Comparison of test-taking behavior scales by cluster

Variable	Cluster 1 *moderate* ($n = 741$, 39.2%) M (SD)	Cluster 2 *calm risk taker* ($n = 637$, 33.7%) M (SD)	Cluster 3 *test anxious risk averse* ($n = 513$, 27.1%) M (SD)
Risk-taking	19.44 (1.9)	20.30 (2.6)	14.43 (2.2)
Motivation	23.56 (5.0)	18.41 (4.4)	24.91 (4.9)
Test anxiety	22.49 (4.4)	13.38 (3.6)	26.44 (4.8)

Table 3 Comparison of demographic and SweSAT specific characteristics by clusters

Variable	Total	Cluster 1 moderate	Cluster 2 calm risk taker	Cluster 3 test anxious risk averse	p value
Demographic characteristics	N (%)	n (%)	n (%)	n (%)	
Gender (male)	759 (40.1)	246 (33.2)	322 (50.5)	191 (37.2)	.000[a] (2 > 1.2*)
Age by group					.002[a]
20 years	1065 (56.3)	420 (56.5)	364 (57.1)	281 (54.8)	
21–24 years	434 (23.0)	177 (23.9)	123 (19.3)	134 (26.1)	2 < 1.3*
25–29 years	178 (9.4)	65 (8.8)	62 (9.7)	51 (9.9)	
30–39 years	143 (7.6)	58 (7.8)	48 (7.5)	37 (7.2)	
40	71 (3.8)	21 (2.8)	40 (6.3)	10 (1.9)	2 > 1.3*
Educational qualifications					.086[a]
Primary school	17 (0.9)	10 (1.3)	4 (0.6)	3 (0.6)	
Secondary school	1524 (80.6)	609 (82.2)	492 (77.2)	423 (82.5)	
Higher education	328 (17.3)	115 (15.5)	132 (20.7)	81 (15.8)	
Not reported	22 (1.2)	7 (0.9)	9 (1.4)	6 (1.2)	
SweSAT specific characteristics	M (SD)	M (SD)	M (SD)	M (SD)	
Achievement level					
Total raw score	86.56 (22.6)	85.35 (22.7)	90.96 (22.6)	82.84 (21.6)	.000[b] (2 > 1.3)
Quantitative	39.11 (12.7)	38.48 (12.6)	41.00 (13.3)	37.68 (11.7)	.000[b] (2 > 1.3)
Verbal	47.50 (13.4)	46.93 (13.5)	50.03 (13.1)	45.16 (13.2)	.000[b] (2 > 1.3, 1 > 3)
Number of tests completed	N (%)	n (%)	n (%)	n (%)	.000[a]
None	1154 (61.0)	464 (62.6)	426 (66.8)	264 (51.5)	3 < 1.2*
One to two	628 (33.2)	234 (31.6)	196 (30.8)	198 (38.6)	3 > 1.2*
Three to four	109 (5.8)	43 (5.8)	15 (2.4)	51 (9.9)	3 > 1.2*, 1 > 2*
Perceived difficulty	N (%)	n (%)	n (%)	n (%)	
Verbal part					.000[a]
Easy	168 (9.1)	55 (7.6)	86 (13.8)	27 (5.3)	2 > 1.3*
Neither easy nor difficult	984 (53.1)	383 (52.8)	375 (60.2)	226 (44.8)	2 > 1.3*, 1 > 3*
Difficult	702 (37.9)	288 (39.7)	162 (20.2)	252 (49.9)	2 < 1.3*, 1 < 3*
Quantitative part					.000[a]
Easy	102 (5.5)	37 (5.1)	50 (8.1)	15 (3.0)	2 > 3*
Neither easy nor difficult	509 (27.6)	184 (25.4)	206 (33.6)	119 (23.7)	2 > 1.3*
Difficult	1230 (66.8)	504 (69.5)	358 (58.3)	368 (73.3)	2 < 1.3*
	N (Md)	Mean rank (Md)	Mean rank (Md)	Mean rank (Md)	
Well prepared	1880 (2.0)	958.7 (2.0)	918.0 (2.0)	942.2 (2.0)	.356[b]
Should have prepared better	1878 (4.0)	924.1 (4.0)	885.9 (4.0)	1028.8 (4.0)	.000[b] (3 > 1.2)

[a] Chi^2 (The Kruskal-Wallis test was used in the variables regarding preparation)

[b] One-way ANOVA

*Column proportions which differ significantly from each other at the .05 level

Cluster 2, calm risk taker

About 50% of the individuals in this cluster are males, a significantly larger proportion compared to the other two clusters (see Table 3). This cluster is also characterized by having a significantly larger proportion of individuals above 40 years old. This group is further

characterized by having a significantly higher achievement level (i.e., SweSAT result) than the other two clusters. Cluster 2 also has the smallest proportion of repeaters. The test takers in this cluster reported that they perceived both parts of the test (i.e., the verbal part and the quantitative part) as easier to a significantly larger degree when compared to the other two clusters. They reported about the same level of preparation as the other two clusters.

Cluster 3, test anxious risk averse

This cluster is made up by about 62% females, which is a significantly larger proportion when compared to cluster 2 but similar to cluster 1 (see Table 3). The age distribution does not differ significantly from cluster 1, but it does differ from cluster 2. Thus, the majority of the individuals in this cluster were younger than 25 years, and only a small part was above 40 years old. This cluster is also characterized by having the lowest result with regard to the SweSAT score. For the total score, this difference is only significant when compared to cluster 2, but the result in the verbal part of the SweSAT is also significantly weaker when compared to cluster 1. Further, this cluster has a significantly larger proportion of repeaters than clusters 1 and 2. Cluster 3 is also characterized by reporting significantly higher levels of perceived difficulty of the quantitative part, as well as the verbal part of the SweSAT, when compared to the other two clusters. Finally, test takers in this cluster reported to be prepared to the same extent as the other clusters, but reported to a significantly higher degree that they could have prepared better when compared to the other two clusters.

Discussion

The present study aimed at classifying test takers in a high-stakes testing situation into subgroups, or profiles, based on their self-reported risk-taking behavior, test-taking motivation, and test anxiety during the test. The analysis generated three profiles that were labeled as follows: *moderate*, *calm risk taker*, and *test anxious risk averse*. As indicated by the labels, the result from the cluster analyses showed three distinct groups of test takers reporting different patterns of test-taking behavior. When comparing the three profiles in relation to demographic and test-specific characteristics, the results revealed additional differences between the profiles, aiding the understanding of the successful test taker.

From a performance perspective, the most successful test taker profile in this study is the *calm risk taker*. This profile scored significantly higher on the total SweSAT, as well as on the verbal and the quantitavie part, respectivly, when compared to the other two profiles. Compared to the other profiles, the test takers in this profile reported low test anxiety and motivation, but relatively high on risk-taking. The fact that this profile reported low levels of test anxiety and willingness to guess when not knowing the answer is in line with earlier research, and with the tentative assumptions made in this study. Test anxiety has been shown to be negatively related to achievement and risk-taking strategies to be positively related (see, e.g., Cheng et al. 2014; Dodeen 2008). Additional research also suggest that to know and use effective test-taking strategies may reduce test anxiety (see, e.g., Taylor and Walton 1997), which might support the pattern of this profile. However, this profile also score low on motivation in relation to the other two profiles. A high level of motivation is generally seen as important for an optimal performance, which is why this may seem as an unexpected result. On the other hand, as motivation increases, test anxiety also tends to increase (see Wolf and

Smith 1995; Smith and Smith 2002), and as anxiety and motivation have opposite effects on performance, it could be assumed that a "high-enough" level of motivation coupled with a "low-enough" level of anxiety would be the best combination in practice. Also, this cluster indeed scored lower on the motivation scale than the other two clusters, but in absolute terms, they still reported a fair level of motivation for doing their best on the test.

Examining this profile further, it was characterized as somewhat older, with a larger proportion of males compared to the other two profiles, and percieving both parts of the SweSAT as relatively easy even though a large proportion of the test takers in this profile did the test for the first time. These results are also expected and in line with erlier research regarding test-taking behavior. That males report lower levels of test anxiety compared to females has been shown in a number of erlier studies (see for example Cassady and Johnson 2002; Stenlund et al. 2017). Previous studies have also shown that level of test anxiety might be correlated with percieved difficulty (Hong et al. 2006). The *calm risk takers* did not perceive the test as especially difficult compared to the other two profiles, which might explain the low levels of reported test anxiety.

The least successful profile from an achievment level pespecive is cluster 3, the *test anxious risk averse* test taker. This cluster performed significantly worse than the most successful profile (the *calm risk taker*) on both the verbal and the quantitative part of SweSAT and significantly worse on the verbal part compared to profile number 1, the *moderate*. The *test anxious risk averse* reported the highest levels of test anxiety and motivation, but also the lowest levels of risk-taking. These patterns align with findings from previous research showing that high test anxiety and low risk-taking is related to poor performance and that high test anxiety might interfere with use of effective test-taking strategies (Peng et al. 2014). Moreover, it might also support the hypothesis that when motivation increases, test anxiety also increases (Wolf and Smith 1995).

Further, and contrary to our preliminary assumptions, what also characterized this profile is that they have the largest amount of repeaters. These results contradict earlier studies suggesting that test takers might be less test anxious when they are familiar with the test and the test situation (see for example, Szpunar et al. 2013), and that repeated test taking is associated with a higher performance (see for example, Stenlund et al. 2016). This profile also perceived the SweSAT as more difficult than the most successful profile, which is in line with the idea that perceived difficulty is related to test anxiety. However, perceived difficulty and high test anxiety might just as well be a consequence of insufficient knowledge and being poorly prepared (Cassady and Johnson 2002). This profile also reported that they should have prepared better to a significantly higher degree than the other two profiles.

The *moderate* profile also reported relatively high motivation and high test anxiety, but the *moderate* test taker is more successful compared to the *test anxious risk averse* test taker. One key difference between these two profiles is the reported levels of risk-taking, suggesting that this test-taking strategy might indeed be important in high-stakes testing situations. Further, this profile did not have as many repeaters and did not experience the test as difficult as the *test anxious risk averse* test taker. Again, this suggests that perceived difficulty and test anxiety might be related. When compared to the *calm risk taker* profile, the key difference is the reported levels of test anxiety and motivation. The *moderate* group reported higher levels of test anxiety as well as motivation compared to the *calm risk taker*. In sum, the results in this study reveal that the pattern of test-taking behavior observed in the *calm risk taker* profile seems to be the most successful in a high-stakes testing situation, and the test-taking behavior pattern of the *test anxious risk averse* the least successful, leaving the profile *moderate* in-between.

The results in the present study need to be interpreted in the context of some limitations. The study sample is not representative from a statistical significance perspective, as the respondents in

this sample to a larger extent are females and high achievers and are somewhat older, compared to the population. However, because of the large sample size and the small effect sizes when comparing differences in distribution between the sample and the total population, we suggest that the results should be considered as representative of what would have been the case in the population. In addition, because the study is based on between-group analyses, it is not apparent what effect these differences might have on the results. Another problem is the measures used to classify the test takers into clusters. First, the low level of reliability of the scale measuring risk-taking ($\alpha = .52$) might be a threat to the validity of the scale. Still, as the scale has few items, of which some are reversed, it could be argued that the alpha value is tolerable according to Field (2013). Second, the fact that the measures used to form clusters are self-reported instruments. Self-reports might be exposed to response bias, such as self-representation bias (Bradburn et al. 2004). For example to fill out a questionnaire in accordance to what the test taker thinks the assessor expects, or overreporting for socially desirable behaviors. Third, as it is impossible to ask the test takers about their experiences during the test administration, the questionnaire was administrated post hoc. It is difficult to say to what extent the respondent's memory of the test situation is accurate. Still, the majority of the participants answered the questionnaire within a few days from the test administration, indicating that they should remember the test situation relatively well. A suggestion for future research is to, if possible, complement self-reported data with data from measures that are more objective.

Although the present study is exploratory and not without limitations, we believe that it provides new and potentially valuable insights about differences between test takers in terms of how test takers approach a test. The study has revealed information about characteristics of both the successful and the less successful test taker that can contribute to better understanding students' performance within high-stakes testing situations. Findings from the present study need to be corroborated by future studies, but the results suggest that the impact of non-cognitive variables in the test situation may be important to acknowledge to better understand student performance and group differences in performance. A better understanding may also have practical consequences. For example, according to our findings, a combination of low test anxiety and high levels of risk-taking seems beneficial for performance, and this combination is more common in male than in female test takers. This makes it possible to pay more attention to these aspects and take appropriate actions, such as providing suggestions of how test takers should prepare for the test and how test proctors may behave during test administration. An increased awareness of non-cognitive factors that are important for performance in high-stakes test situations may then eventually lead to a more pleasant testing experience for the test taker and more accurate test scores.

References

Alnabhan, M. (2002). An emperical investigation of the effects of three methods of handling guessing and risk taking on the psychometric indices of a test. *Social Behavior and Personality, 30*(7), 645–652.

Baldiga, K. (2014). Gender differences in willingness to guess. *Manag Sci, 60*(2), 434–448.

Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. *J Educ Meas, 28*(1), 23–35.

Bicak, B. (2013). Scale for test preparation and test taking strategies. *Educational Science: Theory and Practice, 13*(1), 279–289.

Bond, L., & Harman, A. E. (1994). Test-taking strategies. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 2, pp. 1073–1077). New York: MacMillian Publishing Company.

Boyer, T. W. (2006). The development of risk-taking: A multi-perspective review. *Dev Rev, 26*(3), 291–345.

Bradburn, N. M., Sudman, S., Blair, E., & Stocking, C. (2004). Question threat and response bias. In M. Bulmer (Ed.), *Questionnaires* (Vol. 3, pp. 47–60). London: SAGE Publications Ltd..

Carter, R., Williams, S., & Silverman, W. K. (2008). Cognitive and emotional facets of test anxiety in African American school children. *Cognit Emot, 22*(3), 539–551.

Cassady, J. C., & Johnson, R. B. (2002). Cognitive test anxiety and academic performance. *Contemp Educ Psychol, 27*, 270–295.

Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET and GEPT. *TESOL Q, 48*(2), 300–330.

Culler, R. E., & Holahan, C. J. (1980). Test anxiety and academic performance: The effects of study-related behaviors. *J Educ Psychol, 72*, 16–20.

DeMars, C. E., Bashkov, B. M., & Socha, A. B. (2013). The role of gender in test-taking motivation under low-stakes conditions. *Research & Practice in Assessment, 8*, 69–82.

Dodeen, H. (2008). Assessing test-taking strategies of university students: Developing a scale and estimating its psychometric indices. *Assessment & Evaluation in Higher Education, 33*(4), 409–419.

Dodeen, H. M., Abdelfattah, F., & Alshumrani, S. (2014). Test-taking skills of secondary students: The relationship with motivation, attitudes, anxiety and attitudes towards tests. *South African Journal of Education, 34*(2), 1–18.

Eklöf, H., & Nyroos, M. (2013). Pupil perceptions of National Tests in terms of perceived importance, invested effort and test anxiety. *Eur J Psychol Educ, 28*, 497–510.

Ellis, A. P. J., & Ryan, A. M. (2003). Race and cognitive-ablity test performance: The mediating effects of test preparation, test-taking strategy use and self-efficacy. *J Appl Soc Psychol, 33*(12), 2607–2629.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4:E utg.)*. London: SAGE Publications Ltd..

Geiger, M. A. (1997). An examination of the relationship between answer changing, testwiseness, and examination performance. *J Exp Educ, 66*(1), 49–60.

Hembree, R. (1988). Correlates, causes, effects, and treatmeant of test anxiety. *Rev Educ Res, 58*(1), 47–77.

Hirschfeld, M., Moore, R. L., & Brown, E. (1995). Exploring the gender-gap on the Gre subject test in economics. *Journal of Economic Education, 26*(1), 3–15.

Hong, E., Sas, M., & Sas, J. C. (2006). Test-taking strategies of high and low mathematics achievers. *J Educ Res, 99*(3), 144–155.

Huntley, C. D., Young, B., Jha, V., & Fisher, P. L. (2016). The efficacy of interventions for test anxiety in university students: A protocol for a systematic review. *Int J Educ Res, 77*, 92–98.

Kim, Y. H., & Goetz, E. T. (1993). Strategic processing of test questions: The test marking responses of college students. *Learn Individ Differ, 5*(3), 211–218.

Knekta, E., & Eklöf, H. (2015). Modeling the test-taking motivation construct through investigation of psychometric properties on an expectancy-value based questionnaire. *Journal for Psychoeducational Assessment, 33*, 662–673. doi:10.1177/0734282914551956.

Naylor, F. D. (1997). Test-taking anxiety and expectancy of performance. In J. P. Keeves (Ed.), *Educational Resarch, methodology, and measurement: An international handbook* (2nd ed., pp. 971–976). Cambridge: Cambridge University Press.

Peng, Y., Hong, E., & Mason, E. (2014). Motivational and cognitive test-taking strategies and their influence on test performance in mathematics. *Educational Research and Evaluation: An International Journal on Theory and Practice, 20*(5), 366–385.

Seipp, B. (1991). Anxiety and academic performance. A meta-analysis of findings. *Anxiety Research, 4*, 27–41.

Smith, L. F., & Smith, J. K. (2002). Relation of test-specific motivation and anxiety to test performance. *Psychol Rep, 91*, 1011–1021.

Stenlund, T., Sundström, A., & Jonsson, B. (2016). Effects of repeated testing on shortand long-term memory performance across different test formats. *Educational Psychology: An International Journal of Experimental Educational Psychology, 36*(10), 1710–1727. doi:10.1080/01443410.2014.953037.

Stenlund, T., Eklöf, H., & Lyrén, P.-E. (2017). Group differences in test-taking behavior: An example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice, 24*(1), 4–20. doi:10.1080/0969594X.2016.1142935.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instr Sci, 26*, 127–140.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 110*(16), 6313–6317.

Taylor, K., & Walton, S. (1997). Co-opting standardized tests in the service of learning. *Phi Delta Kappan, 79*(1), 66–70.

van Barneveld, C. (2007). The effect of examinee motivation on test construction within an IRT framework. *Appl Psychol Meas, 31*(1), 31–46.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educ Assess, 10*(1), 1–17.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurment in Education, 8*(3), 227–242.

Woodley, M. A., te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the brand effect. *Intelligence, 43*, 27–34.

Zeidner, M. (1998). *Test anxiety – The state of the art*. New York: Plenum.

Tova Stenlund. Department of Psychology, Umeå University, SE-901 87 Umeå. Mail: tova.stenlund@umu.se; Web site of department: http://www.psy.umu.se/english/?languageId=1

Current Themes of Research:

Tova Stenlund is assistant professor at the Department of Psychology, Umeå University, Sweden. She has a PhD degree in Educational Measurement and her research interest lie in the field of educational psychology and cognitive psychology, with a particular focus on validity in test and assessment results, test-taking behavior, and effects of repeated testing on memory and learning.

Most relevant publications in the field of Psychology of Education (max: 5):

Stenlund, T., Eklöf, H., & Lyrén, P., (2016). Group Differences in Test-Taking Strategies in a High Stake Test Situation. *Assessment in Education: Principles, Policy and Practice*. doi: 10.1080/0969594X.2016.1142935

Stenlund, T., Jönsson, F. U., & Jonsson, B. (2016). Group Discussions and Test Enhanced Learning: Individual Learning Outcomes and Personality Characteristics. *Educational Psychology: An International Journal of Experimental Educational Psychology.* doi:10.1080/01443410.2016.1143087

Stenlund, T., Sundström, A., & Jonsson, B., (2014) Effects of repeated testing on short- and long-term memory performance for different test format. *Educational Psychology: An International Journal of Experimental Educational Psychology,* doi:10.1080/01443410.2014.953037.

Stenlund, T. (2013). Validity in admission decisions related to assessment of prior learning in higher education. *Assessment and Evaluation in Higher education. 38*(1), 1–15

Stenlund, T. (2012). Threats to the valid use of assessment of prior learning in higher education: Claimants' experiences of the assessment process. *Assessment in Education: Principles, Policy and Practice.* 19(2), 177–192.

Per-Erik Lyrén. Department of Applied Educational Science, Umeå University, SE-901 87 Umeå. Mail: per-erik.lyren@umu.se; Web site of department: http://www.edusci.umu.se/english/?languageId=1

Current Themes of Research:

Per-Erik Lyrén is assistant professor at the Department of Applied Educational Science, Umeå University, Sweden. He has a PhD degree in Educational Measurement and his research interests lie in the fields of college admissions testing and measurement of patient-reported outcomes, with a particular focus on score reporting, validation, equating, responsiveness and test-takers' perspectives on testing.

Most relevant publications in the field of Psychology of Education (max: 5):

Stenlund, T., Eklöf, H., & Lyrén, P.-E., (2016). Group Differences in Test-Taking Strategies in a High Stake Test Situation. *Assessment in Education: Principles, Policy and Practice*. doi: 10.1080/0969594X.2016.1142935

Wedman, J., & Lyrén, P.-E. (2015). Methods for examining the psychometric quality of subscores: A review and application. *Practical Assessment, Research & Evaluation, 20*(21). Available at http://pareonline.net/getvn.asp?v=20&n=21

Lyrén, P.-E. (2009). Reporting subscores for college admission tests. *Practical Assessment, Research & Evaluation*, *14*(4). Available at http://pareonline.net/getvn.asp?v=14&n=4

Hanna Eklöf. Department of Applied Educational Science, Umeå University, SE-901 87 Umeå. Mail: hanna.eklof@umu.se; Web site of department: http://www.edusci.umu.se/english/?languageId=1

Current Themes of Research:

Hanna Eklöf is associate professor at the Department of Applied Educational Science, Umeå University, Sweden. Her research interests include the psychology of test-taking, international comparative studies and

measurement quality issues. She currently is the leader of a research project funded by the Swedish Research Council, focusing on test-takers perceptions of tests, test-taking effort and test-taking strategies and the relationship between non-cognitive variables and test performance.

Most relevant publications in the field of Psychology of Education (max: 5):

Stenlund, T., Eklöf, H., & Lyrén, P.-E., (2016). Group Differences in Test-Taking Strategies in a High Stake Test Situation. *Assessment in Education: Principles, Policy and Practice*. doi: 10.1080/0969594X.2016.1142935

Knekta, E., & Eklöf, H. (2015). Modeling the test-taking motivation construct through investigation of psychometric properties on an expectancy-value based questionnaire. *Journal of Psychoeducational Assessment, 33,* 662–673.

Nyroos, M., Jonsson, B., Korhonen, J., & Eklöf, H. (2015). Children's mathematical achievement and its relation to working memory, test anxiety, self-regulation: A person-centered approach. *Education Inquiry, 6*, 73–97.

Eklöf, H., Japelj, B., & Grønmo, L. S. (2014). A cross-national comparison of reported effort and mathematics performance in TIMSS Advanced 2008. *Applied Measurement in Education, 27,* 31–45.

Eklöf, H., & Nyroos, M. (2013). Pupil Perceptions of National Tests in Terms of Perceived Importance, Invested Effort and Test Anxiety. *European Journal of Psychology of Education, 28*, 497–510.