Psychology Science Quarterly, Volume 50, 2008 (1), p. 75-87

The effect of success probability on test economy and self-confidence in computerized adaptive tests

JOACHIM HÄUSLER[1] & MARKUS SOMMER[2]

Abstract

Recent research on the psychological effects of different design decisions in computerized adaptive tests indicates that the maximum-information item selection rule fails to optimize respondents' test-taking motivation. While several recent studies have investigated psychological reactions to computerized adaptive tests using a consistently higher base success rate, little research has so far been conducted on the psychometric (primarily test reliability and bias) and psychological effects (e.g. test-taking motivation, self-confidence) of using mixtures of highly informative (p = .50) and easier items (p = .80) in the item selection process. The present paper thus compares these modifications to item selection with a classical maximum-information algorithm. In a simulation study the effect of the different item selection algorithms on measurement precision and bias in the person parameter estimates is evaluated. To do so, the item pool of the Lexical Knowledge Test, measuring crystallized intelligence and self-confidence, is used. The study indicated that modifications using base success probabilities over p = .70 lead to reduced measurement accuracy and - more seriously - a bias in the person parameter estimates for higher ability respondents. However, this was not the case for the motivator item algorithm, occasionally administering easier items as well. The second study (n = 191) thus compared the unmodified maximum-information algorithm with two motivator item algorithms, which differed with regard to the percentage of motivator items presented. The results indicate that respondents yield higher self-confidence estimates under the motivator item conditions. Furthermore, the three conditions did not differ from each other with regard to the total test duration. It can be concluded that a small number of easier motivator items is sufficient to preserve test-taking motivation throughout the test without a loss of test economy.

Key words: adaptive testing, test economy, success probability, test taking motivation

[1] Address correspondence to: Joachim Häusler, Hyrtlstraße 45, 2340 Mödling, Austria; Phone: +43-2236-4231529, email: haeusler@schuhfried.at

[2] Markus Sommer, Schuhfried GmbH; Phone: +43-2236-4231516, email: sommer@schuhfried.at

Introduction

As a result of continuing improvements in computer technology and psychometrics, computerized adaptive tests are becoming more common in high-stake assessment. In contrast to conventional linear tests with fixed item sequence, the item administration in a computerized adaptive test is dynamically adjusted to the estimate of the respondent's ability. This approach requires a sufficiently large item pool calibrated by means of 1-PL model (Rasch, 1980), the 2-PL and 3-PL model (Birnbaum, 1968) or derivatives thereof (Difficulty plus Guessing Parameter Model; Kubinger & Draxler, 2006). To date most item selection algorithms are based on the maximum-information principle. This means that the algorithm searches the available pool for items that are expected to most successfully minimize the standard error of measurement, by maximizing the Fisher information function (Schervish, 1995; Timminga & Adema, 1995). Within the framework of the 1-PL model this is achieved when the difficulty of item σ_i corresponds best to the ability of the respondent ξ_j. In this case the success probability for the chosen item is $p = .50$. After the chosen item has been administered, the person parameter estimation algorithm re-estimates the respondent's latent person parameter and passes this information to the item selection algorithm so that further items can be selected. This process continues until certain stopping criteria (e.g. number of items; standard error of measurement; total test duration) have been reached.

The typical argument commonly put forward in favour of computerized adaptive tests stresses the fact that computerized adaptive tests generate more information per item and therefore reach a given standard error of measurement earlier compared to conventional linear fixed-item tests (e.g. Hornke, 1993; Sands, Waters & McBride, 1997). Furthermore, several authors have assumed that this item selection algorithm would create a challenging and optimally motivating assessment situation in which subjects feel neither over- nor underchallenged because they are not forced to work on items that are either too hard or too easy for them. However, work on achievement motivation psychology sheds doubt on this assumption. Koestner and McClelland (1990) postulated that moderate task difficulties are generally preferred, where success can be achieved by investing enough effort. In line with this argument Andrich (1995) noted that a success probability of 50 % might be too low to maintain respondents' test-taking motivation in a computerized adaptive test because they will only be able to succeed in approximately 50 % of the items administered to them. This is typically less than the success rate respondents are used to in tests with fixed item sequence.

In constructing a computerized adaptive test several design decisions have to be made that could potentially affect the psychometric characteristics and efficiency of the test as well as respondents' test-taking motivation. For instance, a test developer can choose to start the test with an item of low, intermediate or high difficulty. Research by Lunz and Bergstrom (1994) indicates that this initial item difficulty choice does not affect respondents' performance on a computerized adaptive test. However, some psychometricians (cf. Bergstrom & Lunz, 1999; Mills, 1999; Mills & Stocking, 1996) believe that easier start items enhance self-confidence in one's own ability to master the task at hand, which is known to have a positive effect on respondents' emotional and motivational reactions to the task (c.f. Helmke, 1992). In order to manipulate respondents' self confidence one could systematically alter the difficulty of all subsequent items presented in a computerized adaptive test. From a psychometric point of view this necessarily results in a loss of test information. More precisely, the

information generated decreases as the difference between the ability parameter (ξ_j) and the item difficulty parameter (σ_i) increases (cf. Hambleton, Swaminathan & Rogers, 1991). However, the shape of the information function is fairly flat, indicating that there may be very little loss of information as long as the success probability deviates within reasonable limits from the psychometrically ideal of $p = .50$. In line with this assumption, research by Bergstrom et al. (1992) as well as Tonidandel and Quiñones (2000) indicates that easier computerized adaptive tests targeted at a success probability of $p = .70$ only slightly increase the number of items required to reach a certain level of measurement precision, defined by the standard error of measurement. These results were replicated and extended by Ponsoda, Olea, Rodriguez and Revuelta (1999) and Tonidandel, Quiñones and Adams (2002), who also investigated respondents' reactions to these easier versions of their computerized adaptive tests. These researchers demonstrated that these easier computerized adaptive tests are much more favorable in terms of respondents' test-taking motivation.

The number of items needed to reach a certain precision of measurement is often used as an index of efficiency since it is assumed to be linked to the total test duration required (cf. Wainer, 1993). However, recent research on response latencies in computerized adaptive tests sheds doubt on the relationship between the number of items required to reach a certain measurement precision and the total time needed to complete the computerized adaptive test. For instance, Hornke (1995, 2000) demonstrated that failed items require more time than correct ones. This result has been further validated by several other authors using both computerized adaptive tests and linear fixed-item tests and is commonly known as the "false > correct phenomenon" (e.g. Beckmann, 2000; Hornke, 1995, 2000; Klinck, 2006; Rammsayer, 1999; Preckel & Freund, 2005; Zahaya & Tuvia, 1998). This line of research indicates that the response time is directly related to the success probability. It thus seems reasonable to assume that the efficiency of computerized adaptive testing indicated by the decreased number of items to be administered in comparison to a linear fixed-item test is not necessarily reflected in the test duration. Initial indications of the validity of this assumption are contained in the work of Wild (1989); she noted that response time per item may increase significantly, thus offsetting or even outweighing the saving in the number of items that need to be presented. Häusler (2006) recently presented an algorithm that reduces the total test duration by adapting the success probability for each item to the individual working style of the respondent. Because testing time is often scarce in high-stake selection contexts, even slight gains in test duration might be considered to be worthwhile.

Methods

The following two studies are based on a modified version of the Lexical Knowledge Test (Wagner-Menghin, 2007). This computerized adaptive test was chosen because it is designed as a multifunctional test (Wagner-Menghin, 2006). This means that the test simultaneously measures working style (here: self-confidence) in addition to a latent ability trait (here: crystallized intelligence).

The respondent is first presented with a word and asked to indicate whether he could provide a definition of it. Afterwards the respondent receives an incomplete definition of that word and has to select two out of eight answer alternatives to complete the definition given. An example of the item layout of the Lexical Knowledge Test is given in Figure 1.

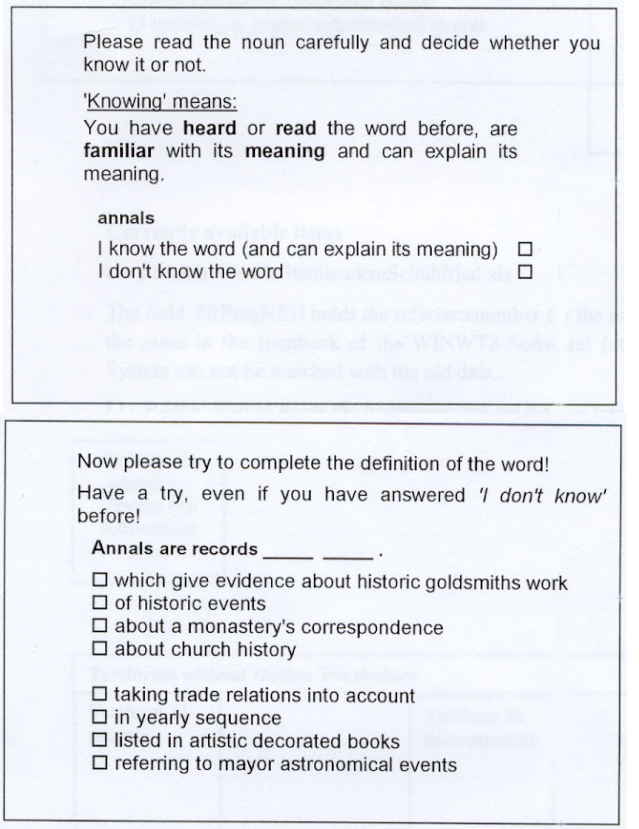

Figure 1:
Instruction items of the Lexical Knowledge Test. Each item is divided into two sections. First the respondent is asked if he thinks he will be able to complete a definition for a specific term (top), then the actual item is presented to him (bottom)

Based on the respondents' judgment of their ability to provide a definition for the words and the actual correctness of the answer, an ability parameter and a working style parameter are calculated. The person parameter for the ability is estimated based on the correctness vector of the responses and the 1PL-item difficulty parameters for the ability test items. In order to estimate the self-confidence person parameter we used the 1PL-item difficulty parameters of the actual ability test items in conjunction with the respondents' judgment of their ability to provide a definition. The benefit of this approach is that the self-confidence person parameter and the actual ability parameter are measured on a comparable scale and the self-confidence indeed reflects the confidence of the respondents in their own ability. Study 1 deals with the psychometric effects of different approaches of making an adaptive test appear easier on test reliability and bias, while Study 2 covers effects on the respondents' self-confidence.

Study 1

The first study examines the psychometric effects caused by systematic manipulation of item success probabilities. Using simulation studies, different methods of making a computerized adaptive test appear easier and therefore presumably more motivating for the respondent are evaluated. For a test modification to be considered reasonable, it should increase the success probability as much as possible without causing a loss in precision of measurement.

Method

Simulation studies are performed with several modifications of the adaptive algorithm based on the item pool of the Lexical Knowledge Test (Wagner-Menghin, 2005). The item pool consists of $k = 126$ items which were calibrated (cf. Wagner-Menghin, 1999, ch. 4.5, 5.3) using the 1-PL model (Rasch, 1980). The difficulty parameters range in the interval [-3.21; 3.88]. Their distribution is approximately normal (Kolmogorov-Smirnoff $Z = .569$; $p = .903$). For each simulation run $n = 20000$ simulees with randomized ability parameter (uniformly distributed between -2.5 and 2.5) were generated. During the starting phase of the adaptive test, depending on success or failure in the first item, the most difficult or most easy item of the pool is presented as the second item. This starting phase design has the advantage that the parameter estimation algorithm can be applied very early – usually after the second item presented. The adaptive algorithm - based on a joint-maximum-likelihood person parameter estimation (JML) - was applied successfully and test reliability after a fixed test length of 20 items was evaluated. The default computerized adaptive setting ($p = .50$), several test modifications using increased base success probabilities ($p = .60$, $p = .70$, $p = .80$) and several test modifications randomly administering a certain percentage (25 %, 50 %) of significantly easier items ($p = .80$) are compared.

Results

The results of the simulation studies show that simply increasing the base success probability of a computerized adaptive test entails some risk. For the item pool of the Lexical Knowledge Test this means that simulees with a very high parameter value of the latent trait suffer a loss in test reliability (see Figure 2).

From Table 1 it can be concluded that this loss in test reliability is a side effect of a parameter estimation bias such that the ability of simulees with a high value on the latent trait is underestimated. The reason for this effect is that after some untypical – but nevertheless possible – cases of "bad luck" during the initial items, the adaptive algorithm using very easy items is unable to collect enough positive information to cause a sufficiently fast increase in the person parameter estimate. Thus at the end of the fixed-length test the respondent's latent trait is still underestimated. If motivator items are used this effect is still visible but to a considerably smaller extent. The reason for this lies in the fact that between the motivator items there are still enough items yielding information that can cause a significant increase in the person parameter estimate.

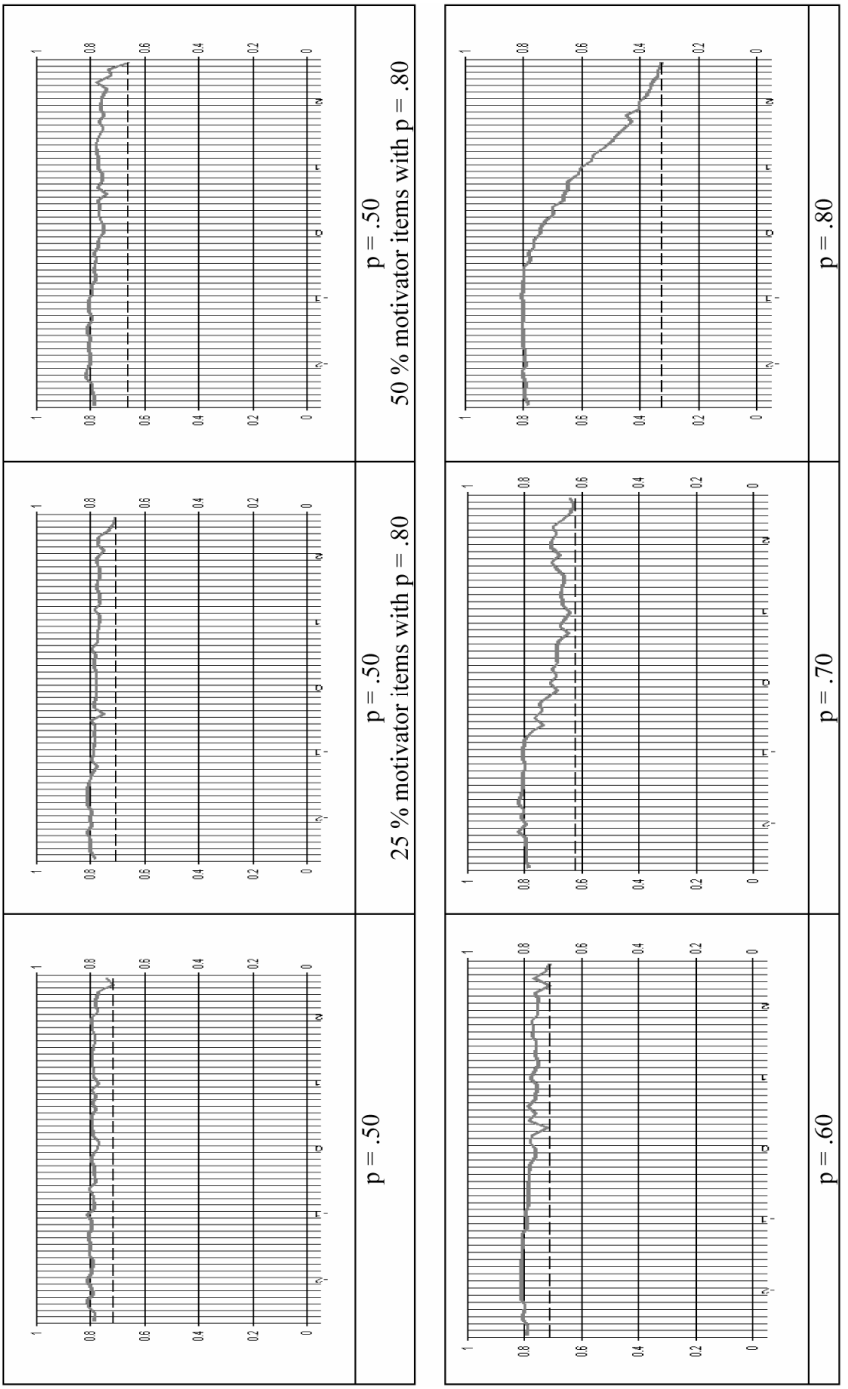

Figure 2:
Simulation of test reliability (y-axis) as a function of the respondent's true score (x-axis) on the latent trait for different settings of the adaptive algorithm

Table 1:
Reliability and bias of the ability parameter estimate using different adaptive algorithms for the Lexical Knowledge Test item pool. Increased success probabilities for all items lead to an underestimation of the ability parameter for respondents with a latent skill in the range of $\xi \geq 2$. This does not happen if items with increased success probability are only administered in an intermittent way

Test form	Reliability after 20 items	Bias after 20 items
p = .50	.801	-.006
p = .60	.771	-.080
p = .70	.723	-.136
p = .80	.624	-.293
p = .50, 25 % motivators (p = .80)	.779	-.019
p = .50, 50 % motivators (p = .80)	.775	-.073

It can thus be concluded that an increase in the base success probability may – depending on the layout of the item pool – cause a bias for high ability respondents by underestimating their ability parameter. The mode of intermittent administration of very easy motivator items does only cause this effect to a very moderate extent.

Study 2

Following the examination of the psychometric effects that intermittent motivator items impose on a computerized adaptive test, the second study deals with the motivational effects of intermittent motivator items. During a test session the respondents' self-confidence is evaluated. To be considered reasonable, a test modification should as far as possible prevent any loss of self-confidence, which could at least for some respondents lead to a decrease in test-taking motivation - during the test session while not interfering with the actual measurement of ability or significantly increasing the test duration.

Method

The respondents were randomized to one of three test forms (0 % motivators being the default maximum-information algorithm, 25 % motivator items, 50 % motivator items). The computerized adaptive test was stopped at a SEM $\leq .44$, which corresponds to a test reliability of α=.83, for all three experimental conditions. Using the multi-functionality of the Lexical Knowledge Test, a working-style parameter is estimated in addition to the ability parameter based on respondents' self-reported belief as to whether or not they will be able to solve the item. The ability parameter and the self-confidence parameter were estimated by means of a joint-maximum-likelihood person parameter estimation (JML). In contrast to the original version of Lexical Knowledge Test, the calculation of the working-style parameter was modified specifically for this study and is thus not available in the custom release of the

test. The intention was to measure respondents' confidence in their own ability. This was done by applying the self-rating response vector to the item difficulty parameters to estimate the respondent's confidence in her or his own ability. Thus the modified self-confidence parameter is an estimate of which ability parameter the respondent would have had, if he had solved all items correctly of which, he thought he could do it.

The number of items required to reach a certain measurement precision, the total test duration as well as the ability and self-confidence person parameter estimates were recorded as dependent variables.

Sample

The sample consists of 84 (44.0 %) male and 107 (56.0 %) female respondents aged between 18 and 50 years. The median age was 33 years with standard deviation of 12 years. A total of 2 (1 %) respondents had not so far completed primary school or had only completed special education (EU educational level 1), 34 (18 %) respondents had completed primary school or basic secondary school but without completing vocational training (EU educational level 2), 67 (35 %) respondents had completed vocational training (EU educational level 3), 65 (34 %) respondents had a qualification at university entrance level (EU educational level 4) and 23 (12 %) respondents had a university degree (EU educational level 5). The three experimental conditions did not differ with regard to educational level ($\chi^2 = 6.525$; $df = 8$; $p = .589$), gender ($\chi^2 = 2.428$; $df = 2$; $p = .297$) or age ($\chi^2 = 3.204$; $df = 2$; $p = .201$).

Results

No significant effects of the three experimental conditions on the ability parameter estimate were found ($F = .890$; $df = 2$; $p = .413$). Since respondents were allocated to the test conditions on a random basis, it can be concluded that the use of motivators does not bias the test result. However, for the self-confidence person parameter a significant difference between the three experimental conditions was observed ($F = 21.1$; $df = 2$; $p < .001$; $\eta^2 = .204$).

As Figure 3 shows, the self-confidence person parameter is higher the more motivators are presented. As can be seen from Table 2, the Scheffé post-hoc test for pairwise comparisons between the three experimental conditions indicates that the two experimental conditions using motivator items differ significantly from the experimental condition using no motivator items. However, the two experimental conditions using motivator items did not differ significantly from each other with regard to the self-confidence person parameter.

If the testing session is considered to be a continuous process, it could be assumed that there is a change in self-confidence as a reaction to the success experienced during the test. From a theoretical point of view one would assume that self-confidence generally declines throughout the test. However, due to differences in the success probabilities between the three experimental conditions one could assume that the decline in self-confidence differs between the three experimental conditions. In order to evaluate this hypothesis separate estimates of respondents' self-confidence were calculated for the first and second halves of the test. These two parameters were subjected to a two-factorial analysis of variance with

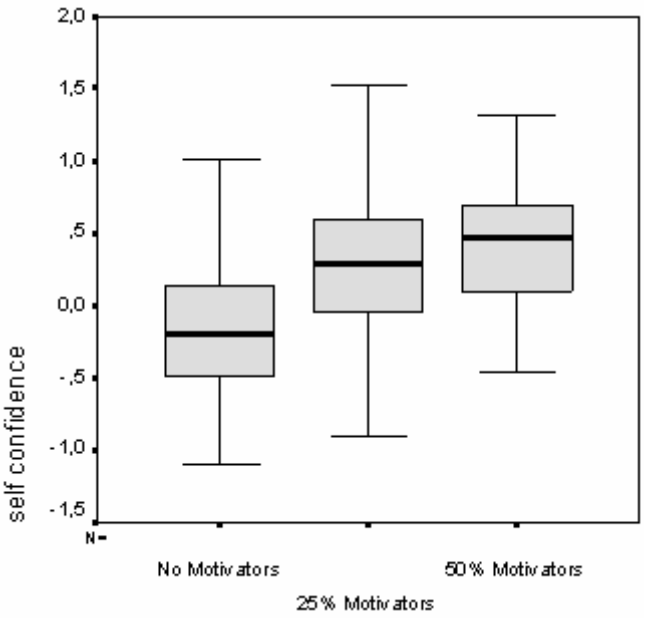

Figure 3:
Box plot of the distribution of the self-confidence person parameter in the three testing conditions

Table 2:
Scheffé post-hoc test for the three experimental conditions: The two test forms using motivator items significantly differ with regard to the person parameter estimate for "self-confidence" from the test form without motivator items. Between the two test forms using motivator items no significant difference in the self-confidence person parameter estimate can be observed

Pair of test conditions	Mean difference	p
No motivators ⇔ 25 % motivators	-.41	< .001
No motivators ⇔ 50 % motivators	-.55	< .001
25% motivators ⇔ 50 % motivators	-.13	.239

one between-subject factor representing the three experimental conditions and one within-subject factor representing respondents' self-confidence parameters for the first and second half of the items administered.

As can be seen from Table 3, there is a significant main effect of the within-subject factor "time" accounting for 38.2 % of the variance and a significant main effect of the between-subject factor "experimental condition" accounting for 17.6 % of the variance. Most importantly, the interaction effect also reached statistical significance and accounts for 3.6 % of the variance in the self-confidence estimates. The results thus indicate that the decrease in self-confidence is sharper for the test form without motivator items.

Table 3:
Analysis of variance results for the factors "time" (first versus second half of the test), "group" (0 %, 25 %, 50 % motivators) and their interaction term. Significant effects on the dependent variable "self-confidence" can be found for all three effects

Effect	F	df	p	η^2
Time	116.0	1	< .001	0.382
Group	20.1	2	< .001	0.176
Time x Group	3.5	2	.032	0.036

As for test economy, the use of motivators did not lead to an increase in test duration (F = 1.574; df = 2; p = .210) even though test length in terms of items to be administered increased significantly due to the use of motivator items (F = 3.778; df = 2; p = .025; η^2 = .039). Taken together the results indicate that the increase in test length can be compensated by a decrease in item response time. Therefore the use of motivator items does not result in a loss in test economy.

Discussion

Even though numerous studies have examined the psychometric effects of different design characteristics of computerized adaptive tests, few studies have investigated the psychological effects of these characteristics. This is surprising, since researchers recognize that selection procedures with comparable validities can have different effects on job applicants' reactions in relation to important organizational outcomes (Smith, Millsap, Stoffey, Reilly, & Pearlman, 1996). Based on the thesis that a success probability of 50 % might be too low to maintain the interest and achievement motivation of respondents, the studies reported in this paper provide further support for the argument that computerized adaptive testing using classic maximum-information item selection algorithms does not always produce an increase in test economy or test reasonableness.

Regarding test reasonableness, the results obtained in Study 2 indicate that, even though self- confidence decreases during a computerized adaptive testing session, this effect was less pronounced for versions of the computerized adaptive tests using intermittent easier motivator items. In consequence, respondents' self-confidence turned out to be significantly lower in the version using a classic maximum-information item selection algorithm than in the two experimental versions using intermittent easier motivator items. This result is in line with previous studies investigating the motivational effects of easier items in computerized adaptive tests. However, while these previous studies found beneficial motivational effects when the item selection algorithm deviated from the maximum-information algorithm by tending to select easier items ($p \geq .70$), the present study obtained similar motivational effects using an item selection algorithm that presented considerably easier items (p = .80) for 25 % or 50 % of the items selected until a certain measurement accuracy was reached. Furthermore, since these two experimental versions did not differ from each other with regard to the beneficial motivational effect, one can conclude that even a small number of easier motivator items might be sufficient to increase the self-confidence of the testees compared to a

classical maximum-information item selection algorithm. This finding is also of relevance in the light of the results obtained in the simulation study conducted for this article. Increasing the base success probability for each single item may result in hazardous effects based on specific characteristics of the item pool. In general an item pool size of 100 to 150 items with uniform – or, even worse, normal - distribution of the item difficulty estimates seems inappropriately small if a test developer intends to apply success probabilities of $p > .60$. In contrast, administering easier items intermittently has two remarkable advantages: (1) the requirements with regard to the item pool are less and (2) it is possible to administer noticeably easier items (e.g. $p \geq .80$) to maintain respondents' test-taking motivation.

With regard to the test economy of computerized adaptive tests, the results obtained in the studies indicate that the administration of easier items will increase the number of items that need to be administered to reach a certain measurement precision. However, the increase in the number of items administered seems to be offset by lower response latencies for the easier items as assumed by Wild (1989). This result is in line with studies on the "false > correct phenomenon", indicating that in general easier items require less time to solve.

In sum, the results reported in this paper indicate that deviating from the maximum-information principle by applying motivator items seems to yield improvements in terms of test reasonableness - as indicated by increased self-confidence person parameters - without suffering any loss in terms of test economy - as indicated by the absence of an increase in the total test duration. Computerized adaptive testing should thus focus less on the opportunity for presenting optimally informative items and instead take possible motivational aspects of test-taking into account. However, the authors acknowledge that the results obtained in this article should be replicated using computerized adaptive tests with different item pool characteristics to further investigate the generalizability of our results.

References

Andrich, D. (1995). Review of the book Computerized-adaptive testing: A primer. *Psychometrika*, *4*, 615-620.

Beckmann, J.F. (2000). Differentielle Latenzzeiteffekte [Differential latencies]. *Diagnostica*, *46*, 124-129.

Bergstrom, B.A.; & Lunz, M.E. (1999). CAT for certification and licensure. In F. Drasgow & J. B. Olsen (Eds.), *Innovations in computerized assessment* (pp. 67-91). Mahwah: Erlbaum.

Bergstrom, B.A.; Lunz, M.E.; & Gershon, R.C. (1992). Altering the level of difficulty in computer adaptive testing. *Applied Measurement in Education*, *5*, 137-149.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (eds.), *Statistical theories of mental test scores* (pp.395-479). Reading: Addison-Wesley.

Hambleton, R.K.; Swaminathan, H.; & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage.

Häusler, J. (2006). Adaptive Success Control in Computerized-Adaptive Testing. *Psychology Science*, *48*, 436-450.

Helmke, A. (1992). *Selbstvertrauen und schulische Leistung*. [Self confidence and achievement in school]. Göttingen: Hogrefe.

Hornke, L.F. (1993). Mögliche Einspareffekte beim computergestützten Testen [Possible efficiency effects of computerized assessment]. *Diagnostica*, *39*, 109-119.

Hornke, L.F. (1995). Item times in computerized testing – A new differential information. *European Journal of Psychological Assessment*, *11*, 108-109.

Hornke, L.F. (2000). Item response times in computerized-adaptive tests. *Psicológica*, *21*, 175-189.

Klinck, D. (2006). *Itembearbeitungszeiten beim computergestützten Testen: Antwortlatenzen bei richtigen und falschen Lösungen*. [Response latencies in computerized testing: latencies of correct and incorrect responses]. Paper presented at the 45. Kongress der Deutschen Gesellschaft für Psychologie (17. – 21. September). Nuremberg: Germany

Koestner, R.; & McClelland, D.C. (1990). Perspectives on competence motivation. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 527 548). New York: Guilford Press.

Kubinger, K.D.; & Draxler, C. (2006). A comparison of the Rasch model and constrained item response theory models for pertinent psychological test data. In M. von Davier & C.H. Carstensen (Eds.), *Multivariate and Mixture Distribution Rasch Models - Extensions and Applications* (pp. 295-312). New York: Springer.

Lunz, M.E.; & Bergstrom, B.A. (1994). An empirical study on computerized-adaptive test administration conditions. *Journal of Educational Measurement*, *31*, 251-263.

Mills, C.N. (1999). Development and introduction of a computer adaptive graduate records examination test. In F. Drasgow & J.B. Olsen (Eds.), *Innovations in computerized assessment* (pp. 117-135). Mahwah: Erlbaum.

Mills, C.N.; & Stocking, M.L. (1996). Practical issues in large-scale computer adaptive testing. *Applied Measurement in Education*, *9*, 287-304.

Ponsoda, V., Olea, J., Rodriguez, M.S., & Revuelta, J. (1999). The effect of test difficulty manipulation in computerized-adaptive testing and self-adapted testing. *Applied Measurement in Education*, *12*, 167-184.

Preckel, F. & Freund, P.A. (2005). Accuracy, latency and confidence in abstract reasoning: The influence of fear of failure and gender. *Psychology Science*, *47*, 230-245.

Rammsayer, T. (1999). Zum Zeitverhalten beim computergestützten adaptiven Testen: Antwortlatenzen bei richtigen und falschen Lösungen [On the latencies in computerized-adaptive tests: latencies for correct and incorrect solutions]. *Diagnostica*, *45*, 178-183.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Sands, W.A.; Waters, B.K.; & McBride, J.R. (1997). *Computerized-Adaptive Testing. From Inquiry to Operation*. Washington DC: American Psychological Association.

Schervish, M.J. (1995). *Theory of Statistics*. New York: Springer.

Smith, J.W.; Millsap, R.E.; Stoffey, R.W.; Reilly, R.R.; & Pearlman, K. (1996). An experimental test of the influence of selection procedures on fairness perception, attitudes about the organization, and job pursuit intentions. *Journal of Business and Psychology*, *10*, 297-318.

Timminga, E.; & Adema, J.J. (1995). Test Construction from Item Banks. In G.H. Fischer & I.W. Molenar (Eds.), *Rasch Models* (pp. 111-127). New York: Springer.

Tonidandel, S.; & Quiñones, M.A. (2000). Psychological reactions to adaptive testing. *International Journal of Selection and Assessment*, *8*, 7-15.

Tonidandel, S.; Quiñones, M.A.; & Adams, A. (2002). Computer-adaptive testing: The impact of test characteristics on percieved performance and test takers' reactions. *Journal of Applied Psychology*, *87*, 320-332.

Wagner, M. (1999). *Lexikon-Wissen-Test (LEWITE) Leistungstest- und/oder Objektiver Test zur Beurteilung der Realitätsangemessenheit der Selbsteinschätzung*. [The Lexical Knowledge

Test. Ability test and/or objective personality test to assess the confidence in one's ability]. Unpublished Dissertation, University of Vienna, Vienna.

Wagner-Menghin, M. (2005). *Manual Lexikon-Wissen-Test* [Lexical KnowledgeTest]. Mödling: Schuhfried.

Wagner-Menghin, M. (2006). Spezielle Multifunktionalität am Beispiel des Lexikon-Wissen-Test. In T.M. Ortner, R.T. Proyer & K.D. Kubinger (Eds.), *Theorie und Praxis Objektiver Persönlichkeitstests* (pp. 204-209) [Theory and Practice of Objective Personality Tests]. Bern: Huber.

Wagner-Menghin, M. (2007). Conception and Construction of a Rasch-Scaled Measure for Self-Confidence in One's Vocabulary Ability. *Journal of Applied Measurement*, *8*, 35-47.

Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. *Educational Measurement: Issues and Practice*, 15-20.

Wild, B. (1989). Neue Erkenntnisse zur Effizienz des "tailored"-adaptiven Testens. In K.D. Kubinger (Ed.), *Moderne Testtheorie* (pp. 169-186) [Modern Test Theory]. Weinheim: Beltz.

Zahaya, D. & Tuvia, R. (1998). Choice latencies times as determinants of post-decisional confidence. *Acta Psychologica*, *98*, 103-115.