Journal of Business and Media Psychology (2011) 1, 12-21 *available at: www.journal-bmp.de*

Towards Positive Test Takers' Reactions to Cognitive Ability Assessments: Development and Initial Validation of the Reasoning Ability at Work Test

Stefan Krumm, Joachim Hüffmeier, Franziska Dietz,
Andre Findeisen, Christian Dries

Westfaelische-Wilhelms-University Muenster

Abstract

The current diffusion of cognitive ability tests in the field of personnel selection is not reflecting their outstanding predictive validity. In the current study, we presume that this paradox is partly due to the fact that cognitive ability tests provide less face validity as compared to other selection tools. In an effort to overcome the gap between the unequivocal research findings and the current practice in the field, a reasoning test was developed with tasks being embedded in a job related context. Two studies were conducted to examine the psychometric properties and the face validity of this test. Results showed that psychometric properties were mediocre (reliability) to good (validity). In addition, the newly developed test provided higher face validity as compared to a matrix test and showed similar face validity as compared to a test aiming at assessing multiple facets of intelligence.

Keywords: general mental ability, test, intelligence, practitioner-researcher divide

1 Introduction

The assessment of general mental ability (GMA), i. e., intelligence or general cognitive ability including reasoning ability at its core (Carroll, 1993), is one of the major topics in psychological research. As early as in 1921, prominent researchers in the field of assessment discussed the question: How can GMA best be measured (Thorndike, 1921)? To date, literally hundreds of different tests are available that assess GMA or one or more of its subdimensions. The tests' predictive validity for job and training performance was confirmed in several meta-analyses, which aggregated studies from all over the world (Schmidt & Hunter, 1998; Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Kramer, 2009). In many countries, however, GMA tests still show a surprisingly low diffusion in the field of personnel selection (Ryan, McFarland, Baron, & Page, 1999). This paradox between diffusion in the field and predictive validity may be explained by the fact that cognitive ability tests are—from an applicant's perspective—perceived as less favorable than other selection procedures (Hausknecht, Day, & Thomas, 2004). The favorability of selection procedures is mainly determined by their face validity and job relatedness (Hausknecht et al., 2004). Hence, the development of GMA tests with high face validity and job relatedness should result in more favorable applicant reactions. The current study describes the development and initial validation of a reasoning ability test, which provides high face validity.

1.1 The Relevance of GMA at Work

GMA is positively and strongly linked to a variety of achievement related outcomes such as educational achievement (e. g., Kuncel, Hezlett, & Ones, 2001), job training success (cf. Schmidt, 2002), and job performance (e. g., Hunter & Hunter, 1984; Schmidt & Hunter, 1998). These relationships have been shown to generalize across cognitive ability tests, across occupations, across career levels, and across nations (for an overview, see Kuncel, Hezlett, & Ones, 2004). Kuncel et al. (2004) concluded that GMA "has been shown to have important, domain-general relationships with knowledge, learning, and information processing, and the general thesis [...] is that tests of general cognitive ability or "g" are predictive of success in academic and work settings, regardless of the setting for which they were developed" (p. 148).

As a consequence of the robust effect of GMA on job performance, researchers have started to ask themselves, why GMA was related to job performance. To date, there is quite substantial evidence that GMA does not directly predict job performance; rather the prediction of GMA on job performance is mediated through the acquisition of job related knowledge (e. g., Schmidt, Hunter, & Outerbridge, 1986) such that smarter individuals find it easier to acquire new knowledge and adapt to new circumstances and, as a consequence, show better job-related performances.

Besides examining the relationship of GMA and performance indicators on the individual level, some researchers have devoted their attention to performance indicators on the organizational level. Terpstra and Rozell (1993), for example, found that the usage of cognitive ability tests as a selection procedure in the service industry was significantly related to the organizations' annual profit, profit growth, and sales growth (with *r*'s around .50).

In sum, the relevance of GMA at work is undisputed. Initial statements that GMA "reveals little about an individual's potential for further growth" (Gardner, 1983, p. 18) have long been falsified by hundreds of studies supporting the outstanding relevance of GMA for many achievement related outcomes in life.

1.2 GMA Tests' Diffusion in the Field

Although the empirical evidence reported in the academic literature is overwhelming, GMA tests are not amongst the most preferred selection procedures in the field (Ryan et al., 1999). In 2003, only around 30% of the surveyed German organizations used GMA tests in personnel selection (Schuler, Hell, Trapmann, Schaar, & Boramir, 2007). Moreover, Schuler et al. revealed that when organizations chose individuals for highly prestigious jobs, they especially refrained from using cognitive ability tests. Although being more pronounced in Germany than in many other countries, the hesitation to use cognitive ability tests is not unique to German organizations: Ryan et al. (1999) found that organizations world-wide only occasionally apply cognitive ability tests. This mismatch between the academic literature and the current practices in the field is often described as practitioner-researcher divide (e. g., Anderson, Herriot, & Hodgkinson, 2001).

Although a practitioner-researcher divide can be found in many areas (cf. Rynes, Giluk, & Brown, 2007), a gap between practices in the field and research findings is especially severe in personnel selection. On the one hand, organizations are missing the opportunity to select those applicants that are most capable to adapt to new circumstances and to acquire job related knowledge. On the other hand, highly intelligent applicants may be rejected on the basis of less valid selection procedures. Both, from an applicant and an organization perspective, bridging the divide between researchers and practitioners in the domain of personnel selection is a vital endeavor.

1.3 Applicant Reactions to GMA Tests

One way of increasing GMA tests' diffusion in the field—and thereby bridging the practitioner-researcher divide—is to administer those tests that lead to more positive applicant reactions (König, Klehe, Berchtold, & Kleinmann, 2010). König et al. examined six theoretically derived aspects as predictors of selection procedure usage (e. g., costs, legality, anticipated applicant reactions) and revealed that anticipated applicant reactions best predicted the usage of selection procedures. This finding is consistent with studies highlighting that selection procedures function as a preview of the organization and as a marketing tool (Premack & Wanous, 1985). An overview of the potential consequences of negative applicant reactions to selection procedures is provided by Hülsheger and Anderson (2009); these authors name, for example, negative image, impact on consumer behavior, negative work attitudes and performance after being hired, as well as legal implications. Apparently, HR managers prioritize these potential consequences and de-prioritize the predictive validity of the selection procedures.

When focusing on applicant reactions to selection procedures, HR managers are ultimately required to ensure face validity and job relatedness of the applied procedures. Face validity (i. e., perceived test relevance to the test taker, Sackett & Lievens, 2007) and job relatedness were found to be among the most relevant aspects to influence applicant reactions (Hausknecht et al., 2004). However, HR managers are likely to face problems in finding cognitive ability tests with the potential to cause positive applicant reactions. Indeed, Kersting (2008) confirmed that a set of the most frequently used German cognitive ability tests yielded only low to moderate face validity and job relatedness ratings. Hence, we conclude that psychometrically sound cognitive ability tests, which provide high face validity and job relatedness, are rare. The aim of the current study is to develop a GMA test with good psychometric properties targeting managerial and highly qualified staff, which also leads to positive test takers' reactions.

1.4 Development of the Reasoning Ability at Work Test

Reasoning is the best predictor of fluid intelligence; fluid intelligence in turn is most closely associated with GMA (Carroll, 1993). Although Carroll (1993) proposed three types of reasoning factors (deductive reasoning, inductive reasoning, and quantitative reasoning), he also admitted that these factors may be difficult to distinguish as any one task may require more than one reasoning factor (cf. Kyllonen & Christal, 1990). Thus, we aimed at assessing reasoning ability in general, i. e., the ability to deduce rules from given pieces of information and to apply these rules in order to solve the task at hand. Typical reasoning tasks are: number series, syllogisms, and matrices. In these tasks, abstract pieces of information are presented (e. g., numbers, geometrical figures). However, in an effort to avoid negative test takers' reactions, we assessed reasoning ability in tasks consisting of material that is relatively common in administrative and managerial jobs (e. g., short reports, statistics, and graphs). Sample tasks are presented in Figures 1 and 2. Besides the use of job-related material, task development was guided by either one out of two principles. The first principle was to present material that was built according to a logical structure. The last piece of information was missing and had to be added by the test taker. However, to correctly add this last piece of information, test takers had to identify the logical structure inherent to the presented material. This principle is evident in many established reasoning tasks such as in number series. In fact, some of the tasks in the reasoning ability at work test represented number series that were framed in a job-related manner (cf. Figure 1). The second principle was to present information (in form of graphs or short texts) along with four different answer alternatives that represented different conclusions. Only one out of the four alternatives represented a conclusion that could be correctly drawn from the presented graphs or texts (cf. Figure 2). This principle is less often used in established reasoning tests but more often in tests of reading comprehension or tests that require interpreting information (see for example, subtest "Interpreting Information" of the Analysis of Reasoning and Creative Thinking Test, Schuler & Hell, 2005). Please note that each task of the Reasoning Ability at Work Test required only the presented information, no additional knowledge was necessary. In fact, we used fictitious reports, statistics, and graphs; thus, participants were not able to benefit from knowledge about related facts. Depending on the task type (see above), answers had to be given by stating the correct numbers in free answer formats or by either choosing one out of four answer alternatives. The final task set consisted of 20 items in German language. Test duration was limited to 35 minutes.

1.5 Aims of the Current Study

We conducted two subsequent studies aiming at (a) assessing the psychometric properties of the newly developed test, and (b) examining test takers' reactions to the work related reasoning test. We hypothesized that the Reasoning Ability at Work Test would yield good psychometric properties (i. e., good item-scale correlations, high consistency, and first evidence for convergent and discriminant validity). Additionally, we posited that test takers' reactions to the Reasoning Ability at Work Test would be more positive as compared to test takers' reactions to a reasoning test using abstract material.

2 Study 1

2.1 Method

*Participants***.** The 20 items of the Reasoning Ability at Work Test were administered to 92 managers from various divisions of a large German organization from the private business sector. The newly developed test was part of a larger test battery aiming at identifying the candidates' potential for further career advancement. However, promotion decisions were not based on the Reasoning Ability at Work Test. To ensure anonymity, no further data about the sample were collected.

*Materials***.** The 20 items of the Reasoning Ability at Work Test were administered in a paper-and-pencil format. Proctored group sessions were conducted with about 10 participants per session.

*Statistical Analyses***.** We assessed means and standard deviations per each test item. Additionally, the bivariate correlations between each item and the total score of the remaining items (part-whole correction) were calculated. Reliability estimates were made using split-half reliability (odd-even method).

2.2 Results

Means, standard deviations and item-total correlations are provided in Table 1. The items means indicated that most item difficulties were medium, whereas 3 items showed low difficulties (mean scores > .70) and 3 items showed high difficulties (mean scores < .30). Remarkably, only one participant answered item 7 correctly.

The correlations between single items and the total score of the scale were .25 on average. Different items showed coefficients below .30 indicating the homogeneity of the test may require improvement.

On average, participants achieved 9.37 correct answers (*SD* = 3.43, Range = 2 to 18) indicating that the overall test was sufficiently difficult for the intended target group. Split-half reliability was found to be low to moderate (r_{tt} = .68). Selecting only those 10 items that yielded item-total correlations of .25 or above, a split-half reliability of r_{tt} = .70 could be achieved. Spearman-Brown prediction (Brown, 1910; Spearman, 1910) revealed that extending this 10-item version to the original length of 20 items, a reliability of r_{tt} = .82 would be achieved.

Spearman-Brown prediction further revealed that a test length of 40 items would result in a reliability estimate of r_{tt} = .81; a test length of 60 items, which is still not unusual in established reasoning tests (e. g., Kersting, Althoff, & Jäger, 2008), in a reliability estimate of r_{tt} = .86.

3 Study 2

3.1 Method

Participants. The same 20 item version of the test, which was used in Study 1, was administered to 89 students of the Hochschule Fresenius (Cologne, Germany) and the University of Cologne (Germany). The majority of participants were students of psychology (65%). Their mean age was 25 years (Range = 19 to 47); their mean academic training was 5.2 semesters (*SD* = 3.8). A fraction of 44% had gained at least some kind of occupational experience before taking the test (e. g., apprenticeships, internships). Most participants were female (71%).

Materials. **(1)** *Reasoning Ability at Work Test.* The 20 items of the Reasoning Ability at Work Test were administered in a paper-and-pencil format. **(2)** *WILDE Intelligence Test (WIT-2; Kersting et. al., 2008).* The WIT-2 is an established test aiming at assessing the multiple facets of GMA. In order to keep test duration at a minimum, we only administered the reasoning module, which comprises 3 subtests (verbal analogies, number series, folding).

Verbal analogies and number series are standard reasoning tests and, thus are not further described. Folding is a subtest that requires participants to decide which one out of five alternative 3-dimensional figures is resembled by an unfolded 2-dimensional model. The WIT-2 was administered in paper-and-pencil format. Test duration for each subtest was limited following recommendations by the test authors; altogether, the reasoning module test duration was about 35 minutes. The WIT-2 reasoning module provided very good reliability estimates (Cronbach's alpha = .94) as well as convergent and discriminant validity evidence (cf. Kersting et al., 2008). To provide further discriminant validity evidence, we additionally applied a module from the WIT-2 which assesses knowledge in the domain of economy. This module contains 20 items. The authors report a reliability estimate of Cronbach's alpha = .81 (Kersting et al., 2008). **(3)** *AKZEPT-L (Kersting, 2008).* The AKZEPT-L is a 16 item questionnaire aiming at assessing test takers' reactions to achievement tests. This questionnaire was specifically developed for administration after having taken the achievement test in question (i. e., the WIT-2). It comprises four dimensions (measured with four items each): perceived psychometric quality, face validity, perceived opportunity to perform, and perceived strain. These dimensions are derived from recommendations by Gilliland (1993). Reliability estimates are reported to range from .65 to .82 (Cronbach's alpha) across several studies (cf. Kersting, 2008). **(4)** *Self-reported Grades.* The self-reported math and German language grades were assessed as obtained in the final year of secondary education. The math grade served as an indicator for convergent validity evidence, whereas the German language grade was considered an indicator for discriminant validity evidence.

Procedure. Proctored group sessions were conducted with about 5 to 10 participants per session. In a within-subjects design, we administered the two reasoning tests (Reasoning Ability at Work Test and WIT-2). Each one of the two reasoning tests was followed by an assessment of test takers' reactions to this test (as captured with the AKZEPT-L). Participants were assigned to two possible test sequences: (1) Reasoning Ability at Work Test – AKZEPT-L – WIT-2 – AKZEPT-L, or (2) WIT-2 – AKZEPT-L – Reasoning Ability at Work Test – AKZEPT-L. Feedback about their performances was given to those participants who wanted to receive feedback.

Statistical Analyses. Validity evidence was obtained by calculating bivariate correlations. Paired *t* tests were conducted to examine the differences in test takers' reactions to both, the Reasoning Ability at Work Test and the WIT-2. Additionally, we compared the test takers' reactions to the Reasoning Ability at Work Test with test takers' reactions to other GMA tests as reported by Kersting (2008). More specifically, we focused on the Raven test (APM; Raven, Raven, & Court, 1998) as this test exclusively contains matrices and thus differs from tests assessing multiple facets of intelligence. The test takers' reaction in the current study can be compared to reactions reported by Kersting (2008) because Kersting's sample was very similar to our sample (students of psychology; mean age of 25). Moreover, test conditions were similar such that participation was voluntary, the test results were not attached to any consequences, and test takers' reactions were assessed immediately after test conduction.

3.2 Results

Descriptive statistics of the applied measures are presented in Table 2. Notably, the Reasoning Ability at Work Test performances of the student sample did not significantly differ from the managerial sample in Study 1, $t(179)$ = 0.35, *ns.*

Bivariate correlations between the Reasoning Ability at Work Test and validity measures are presented in Table 3. The Reasoning Ability at Work Test was strongly related to the WIT-2 reasoning module. A moderate correlation between the Reasoning Ability at Work Test and knowledge in the domain of economy was observed; however, this correlation is comparable to the relationship between the WIT-2 reasoning module and knowledge in the domain of economy. Thus, these findings provide first convergent and discriminant validity evidence.

Similarly, the Reasoning Ability at Work Test was moderately but significantly related to the self-reported math grade. No positive relationship between the Reasoning Ability at Work Test and the German language grade was observed. In fact, this correlation coefficient was negative and significant, indicating that good German language grades were associated with low achievements in the Reasoning Ability at Work Test. A similar result was found for the correlation between German language grades and the knowledge test. We can only speculate that this might be an artifact of the data.

The comparison of test takers' reactions towards the different tests (cf. Table 4) revealed that—contrary to our hypothesis—the WIT-2 yielded more positive test takers' reaction as compared to the Reasoning Ability at Work Test. This was true for every dimension except the face validity dimension. Please note that the initial development of the Reasoning Ability at Work Test placed special importance on its face validity. Presumably, the newly developed Reasoning Ability at Work Test was—in general—not yet designed well enough to keep up with the established WIT-2 with the only exception being its face validity.

The comparison of the test takers' reaction between the Reasoning Ability at Work Test and the Raven test (APM) as reported by Kersting (2008) revealed a mixed picture. Participants judged the Reasoning Ability at Work Test as more face valid and less strain-imposing. On the other hand, subjects rated the Raven test (APM) better in the dimensions opportunity to perform and psychometric quality.

4 General Discussion

The current research reported the development and initial validation of the Reasoning Ability at Work Test. So far, our results draw a mixed picture. The psychometric properties of the newly developed test were mediocre in view of the tests' reliability. A subset of 10 items was identified that may serve as a starting base for test prolongation to 20 homogenous items. Initial validity evidence confirmed the test's validity. However, the use of statistics, reports, and graphs as test material did not result in the expected test takers' perceptions: The Reasoning Ability at Work Test did not show higher face validity as compared to the WIT-2, an established and up-to-date test containing multiple task types. However, the Reasoning Ability at Work Test exceeded a test comprising only matrix items (Raven test, APM) in terms of face validity.

To date, we are still lacking studies that explicitly examine what test features lead to more positive applicant reactions. The current study presented an initial attempt by assessing reasoning ability in the context of job related tasks. However, we did not study the effects of this design in an experimental setting such that the same tasks were either administered in an abstract way or embedded in job related material. Future studies should consider such designs to obtain more clear-cut results. The design of the current study allows many explanations for the fact that the Reasoning Ability at Work Test was evaluated less positively than the WIT-2 in all but the face validity dimension. Possible reasons might be the use of a student sample instead of a managerial sample in Study 2. Moreover, the student sample mainly consisted of psychology students who might be used to abstract test materials. Our findings also indicate that future revisions of the Reasoning Ability at Work Test should not only emphasize face validity but also perceived psychometric property, opportunity to perform, and perceived strain. The finding that face validity was perceived higher in the Reasoning Ability at Work Test as compared to the Raven test (APM) represents a promising result, which should enthuse researchers to further pursue this approach.

The reliability estimate of the newly developed test was mediocre (r_{tt} = .70). In line with this finding, some correlations between single items and the total score of the scale were below .30 indicating a reduced homogeneity of the test. This might be due to the fact that we intended to create a short reasoning test that nevertheless covered different job-related materials (statistics, charts, short reports) as well as different principles of item development (see Test Development section). Please note that test duration is usually a critical issue for practitioners when assessing the tests' applicability. Hence, our effort to bridge the practitioner-researcher divide also involved keeping test duration at a minimum. However, this resulted in an insufficient number of similar items (i. e., similar in test material and task development principle) to assess their psychometric properties individually. Future test revision should be concerned with developing similar sets of items as well as adaptive test administration to further improve reliability and, at the same time, keep test duration at a minimum.

The Reasoning Ability at Work Test yielded very good validity evidence. Its correlation with the WIT-2 was comparable to correlations between other, well established reasoning tests (cf. Amthauer, Brocke, Liepmann, & Beauducel, 2001). The same is true as concerns the correlation with math grades (cf. Krumm, Ziegler, & Buehner, 2008). Future studies should also assess the predictive validity of the Reasoning Ability at Work Test. It is, however, interesting to note that cognitive ability tests embedded in a job related context, which aim at other than managerial target groups (i. e., apprentices), already provided high validity coefficients for the prediction of training success (cf. Görlich & Schuler, 2007; Schuler & Klingner, 2005).

4.1 Limitations

The assessment of test takers' reactions did not involve the intended target group of the newly developed test. Due to organizational reasons, it was not possible to administer the AKZEPT-L to the cohort of managers in Study 1. However, it seems feasible (mainly for economic reasons) to start a line of research with student samples (cf. Hüffmeier, Krumm, & Hertel, in press). Moreover, the student sample was not totally inexperienced regarding the material used in the test (44% reported some kind of occupational experience). Furthermore, this material usually is featured in student textbooks as well. Notwithstanding, this research should be pursued including managerial samples.

As already mentioned, this first study does not systematically vary tasks such that identical tasks are either embedded in a job related context or not. Thus, the possible conclusions are limited. However, we were able to develop a psychometrically sound test as a basis for further research.

4.2 Conclusion

In sum, the current study contributes to the growing body of research on the practitioner-researcher divide (for a perusal of the divide in a specific domain, see for example Hüffmeier et al., in press). The divide is particularly evident in the domain of cognitive ability testing (Ryan et al., 1999) and we believe that it takes a lot of concerted efforts (both from practitioners and researchers) to at least partially bridge the divide. Developing tests that meet the demands of the field while not ignoring psychometric principles—in our opinion—has the potential to represent such a bridge. Hopefully, we will see more tests in the future, which are mutually developed by researchers and practitioners, thereby ensuring that psychometric tests address the needs of the field.

5 References

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R [Intelligence Structure Test 2000 R].* Göttingen: Hogrefe.

Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner-researcher divide in industrial, work and organizational (IWO) psychology: Where we are now, and where do we go from here? *Journal of Occupational and Organizational Psychology, 74*, 391-411.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences* (2nd ed.). New York: Basic Books.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*, 694-734.

Görlich, Y., & Schuler, H. (2007). Personalentscheidung und Nutzen [Personnel decisions and utility]. In H. Schuler & K. Sonntag (Eds.), *Handbuch der Psychologie: Handbuch der Arbeits- und Organisationspsychologie*. Göttingen: Hogrefe.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639-683.

Hüffmeier, J., Krumm, S., & Hertel, G. (in press). The practitioner-researcher divide in psychological negotiation research: Current state and future perspective. *Negotiation and Conflict Management Research.*

Hülsheger, U. R., & Anderson, N. (2009). Applicant perspectives in selection: Going beyond preference reactions. *International Journal of Selection and Assessment, 17*, 335-345.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests [Acceptance of intelligence and achievement tests]. *Report Psychologie, 33*, 420-433.

Kersting, M., Althoff, K., & Jäger, A. O. (2008). *Wilde-Intelligenztest 2 [Wilde intelligence test 2]*. Göttingen: Hogrefe.

König, C. J., Klehe, U.-C., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment, 18*, 17-27.

Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: Vertiefende und weiterführende Metaanalysen [General mental ability and occupational success in Germany: Further metaanalytic elaborations and amplifications]. *Psychologische Rundschau, 60*, 82-98.

Krumm, S., Ziegler, M., & Bühner, M. (2008). Reasoning and working memory as predictors of school grades. *Learning and Individual Differences, 18*, 248-257.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162-181.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*, 148-161.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*, 389-433.

Premack, S. L., & Wanous, J. P. (1985). A meta-analysis of realistic job preview experiments. *Journal of Applied Psychology, 70*, 706-719.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Advanced progressive matrices*. Oxford, UK: Oxford Psychologists Press.

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359-391.

Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal, 50*, 987-1008.

Sackett, P. R., & Lievens, F. (2007). Personnel selection. *Annual Reviews, 59*, 419-450.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology, 56*, 573-605.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15*, 187-211.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439.

Schuler, H., & Hell, B. (2005). *Analyse des schlussfolgernden und kreativen Denkens (ASK) [Analysis of reasoning and creative thinking]*. Bern: Huber.

Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre [Use of personnel selection instruments in German organizations in the last 20 years]. *Zeitschrift für Personalpsychologie, 6*, 60-70.

Schuler, H., & Klingner, Y. (2005). *Arbeitsprobe zur berufsbezogenen Intelligenz – Büro- und kaufmännische Tätigkeiten (AZUBI-BK) [Work sample test of job-related intelligence for office and business functions]*. Göttingen: Hogrefe.

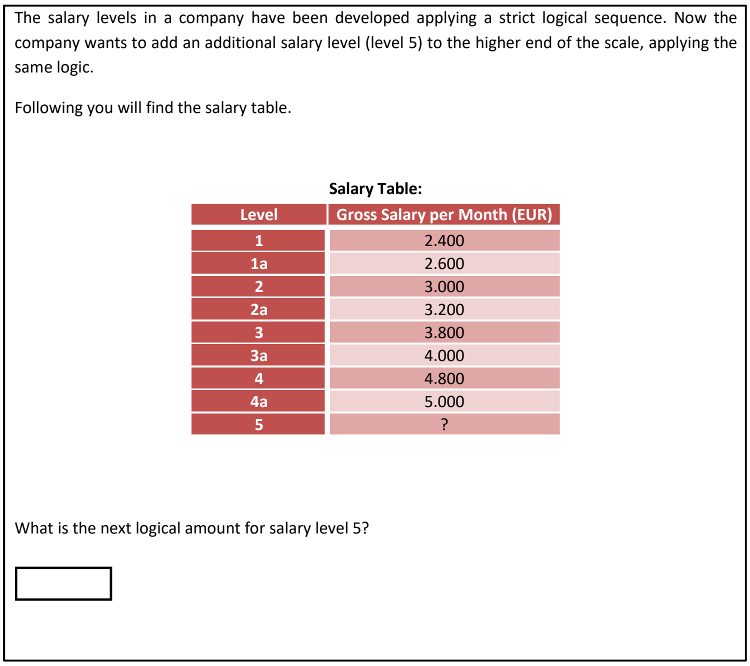
Spearman, C. C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.

Terpstra, D. E., & Rozell, E. J. (1993). The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology, 46*, 27-48.

Thorndike, E. L. (1921). Intelligence and its measurement: A symposium-I. *Journal of Educational Psychology, 12*, 124-127.

Corresponding Author:
Dr. Stefan Krumm
Department of Psychology and Sport Studies
Westfaelische-Wilhelms-University Muenster
Fliednerstrasse 21
D-48149 Muenster
GERMANY
stefankrumm@uni-muenster.de

6 Appendix



The salary levels in a company have been developed applying a strict logical sequence. Now the company wants to add an additional salary level (level 5) to the higher end of the scale, applying the same logic.

Following you will find the salary table.

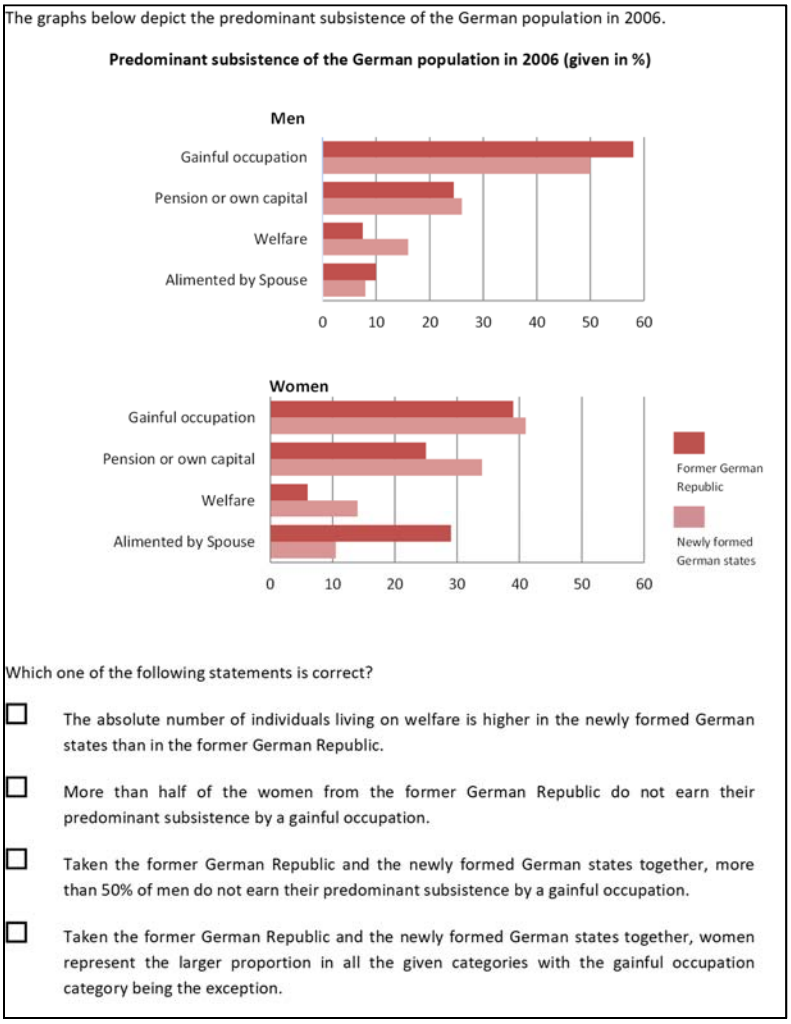
Salary Table:

Level	Gross Salary per Month (EUR)
1	2.400
1a	2.600
2	3.000
2a	3.200
3	3.800
3a	4.000
4	4.800
4a	5.000
5	?

What is the next logical amount for salary level 5?

□

Figure 1. **Sample item 1 of the Reasoning Ability at Work Test.**



The graphs below depict the predominant subsistence of the German population in 2006.

Predominant subsistence of the German population in 2006 (given in %)

Which one of the following statements is correct?

□ The absolute number of individuals living on welfare is higher in the newly formed German states than in the former German Republic.

□ More than half of the women from the former German Republic do not earn their predominant subsistence by a gainful occupation.

□ Taken the former German Republic and the newly formed German states together, more than 50% of men do not earn their predominant subsistence by a gainful occupation.

□ Taken the former German Republic and the newly formed German states together, women represent the larger proportion in all the given categories with the gainful occupation category being the exception.

Figure 2. **Sample item 2 of the Reasoning Ability at Work Test.**

Table 1: Means, Standard Deviations and Item-Total Correlations (Study 1)

Item No.	*M*	*SD*	*rit*	*ri'i*
1	.71	.46	.14	.05
2	.52	.50	.22	.08
3	.52	.50	.09	.04
4	.51	.50	.20	.08
5	.65	.48	.25	.09
6	.53	.50	.27	.10
7	.01	.10	-.01	-.04
8	.78	.41	.23	.08
9	.49	.50	.13	.04
10	.41	.50	.39	.14
11	.67	.47	.34	.13
12	.38	.49	.14	.06
13	.58	.50	.46	.17
14	.33	.47	.31	.11
15	.82	.39	.38	.14
16	.18	.39	.18	.07
17	.26	.44	.25	.09
18	.42	.50	.55	.20
19	.48	.50	.32	.13
20	.11	.31	.11	.04
Total	9.37	3.43		

Note: rit = item – total correlation (part-whole corrected); *ri'i* = average correlation between item and remaining items of the scale

Table 2: Descriptive Statistics of Measures Applied in Study 2

Measure	*M*	*SD*	*rtt*
Reasoning Ability at Work Test (20 items)	9.20	3.11	.64 [3]
AKZEPT-L [1]			
perceived psychometric quality	3.51	.85	.77 [4]
face validity	3.23	.90	.74 [4]
perceived opportunity to perform	4.68	.96	.80 [4]
perceived strain (recoded)	3.25	1.11	.86 [4]
WIT-2 (60 items)	34.24	9.28	.81 [3]
AKZEPT-L [2]			
perceived psychometric quality	3.91	.82	.81 [4]
face validity	3.24	.76	.83 [4]
perceived opportunity to perform	5.46	.60	.66 [4]
perceived strain (recoded)	4.12	1.08	.86 [4]
Math Grade	10.69	2.76	*n.a.*
German Language Grade	11.27	2.65	*n.a.*

Note: rtt = reliability estimate. [1] = AKZEPT-L administered after the Reasoning Ability at Work Test; [2] = AKZEPT-L administered after the WIT-2; [3] = split-half reliability; [4] = Cronbach's alpha

Table 3: Bivariate Correlations

	1	2	3	4
1) Reasoning Ability at Work Test (20 items)	-			
2) WIT-2 (60 items)	.61**	-		
3) Knowledge in the Domain of Economy	.34**	.38**	-	
4) Math Grade	.29**	.32**	.12	-
5) German Language Grade	-.27**	-.08	-.29**	.37**

Note: ** = *p* < .01; * = *p* < .05

Table 4: Comparison of Test Takers' Reactions

	Reasoning Ability at Work Test vs. WIT-2	Reasoning Ability at Work Test vs. Raven test (APM)
AKZEPT-L		
perceived psychometric quality	$t_{(88)}$ = -4.78, *p* < .01	$t_{(153)}$ = -7.61, *p* < .01
face validity	$t_{(88)}$ = -0.22, *ns.*	$t_{(153)}$ = 5.41, *p* < .01
perceived opportunity to perform	$t_{(88)}$ = -7.67, *p* < .01	$t_{(153)}$ = -5.63, *p* < .01
perceived strain (recoded)	$t_{(88)}$ = -6.48, *p* < .01	$t_{(153)}$ = 1.98, *p* < .05

Note: Paired *t* tests were calculated for the comparison between the Reasoning Ability at Work Test and the WIT-2; unpaired *t* tests were calculated for the comparison between the Reasoning Ability at Work Test and the Raven test (APM)