

Please cite as

Peter, J., Leichner, N., Mayer, A.-K., & Krampen, G. (in press). A short test for the assessment of basic knowledge in psychology. *Psychology Learning & Teaching*.

A Short Test for the Assessment of Basic Knowledge in Psychology

Peter, J.¹, Leichner, N.¹, Mayer, A.-K.¹, & Krampen, G.¹

¹ Leibniz Institute for Psychology Information (ZPID)
Universitätsring 15, 54296 Trier, Germany
{peter, leichner, mayer, krampen}@zpid.de

Dr. Johannes Peter was Research Associate at ZPID.

Dr. Nikolas Leichner is Research Associate at ZPID.

Dr. Anne-Kathrin Mayer is Head of the Research Department at ZPID.

Professor Dr. Günter Krampen is the Director of ZPID.

Abstract

This paper reports about the development of a fixed-choice test for the assessment of basic knowledge in psychology for use with undergraduate as well as graduate students. Test content is selected based on a core concepts approach and includes a sample of concepts which are indexed most frequently in common introductory psychology textbooks. In a pilot study ($N = 67$ psychology students), a 21 item-version of the test was developed. Empirical results from two additional studies using psychology students ($N = 173$ and $N = 112$) show that the test can differentiate between students with varying progress in their studies and show a significant correlation of test scores with academic achievement in psychology.

Introduction

The aim of this study is to present a test that can be used to assess basic knowledge in psychology across curricula. Reliable and valid standardized assessment instruments for basic knowledge in psychology are needed for various reasons. Valid and reliable assessment tools are required for several kinds of research efforts. For example, psychology knowledge might be used as an indicator of academic achievement in psychology programs, and can thus be used in studies examining predictors of academic achievement (e.g., Stoloff & Feeney, 2002). Standardized tests might also be used for the evaluation of study programs (American Psychological Association, 2009; Educational Testing Service, 2014a). Furthermore, assessment might improve learning: Several studies show that frequent testing does increase retention for the material tested (see the review by Roediger & Karpicke, 2006).

Based on these considerations, we aimed at creating a relatively short test that can be used as a resource-efficient indicator of basic psychology knowledge. The test described here contains 21 items, and can be completed in around 20 minutes. Therefore the test can be used as part of larger studies which distinguishes it from existing tests.

The research was funded by the German Joint Initiative for Research and Innovation with a grant acquired in the Leibniz Competition 2012.

With regard to test contents, we referred to so-called “core-concepts” in psychology, defined as those concepts cited most frequently by introductory psychology textbooks (Proctor & Williams, 2006). We made sure that every major area of psychology (e.g., cognitive psychology) is covered and that the test includes fundamental concepts (e.g., color perception) as well as more area-specific concepts (e.g., cognitive dissonance reduction) so that it can differentiate between groups with differing levels of knowledge.

Assessment of Basic Knowledge about Psychology

Several standardized tests for the assessment of basic psychology knowledge have been developed for use with English speaking populations. The Major Field Test for Psychology provided by the Educational Testing Service (2014a), for example, consists of 140 multiple-choice items assessing basic concepts, and principles in psychology. According to the test description, “the questions in the psychology test are drawn from the courses of study most commonly offered in undergraduate programs within the broadly defined field of psychology”. The test items cover material from differing fields, like cognition, abnormal psychology, or research methods and can be used for program evaluation. An overall score, as well as subscores related to four fields can be reported. Several studies show that performance on the Major Field Test for Psychology correlates significantly with other measures of academic success, like SAT or college GPA (e.g., Gallagher & Cook, 2013; Stoloff & Feeney, 2002). There is also evidence that more advanced students score higher on the Major Field Test than new students (Pinter, Matchock, Charles, & Balch, 2014).

Educational Testing Service also produces the GRE Subject Test Psychology (GRE-P; Educational Testing Service, 2014b) which is used for admission to several graduate schools in the US. The GRE-P includes around 200 multiple-choice questions and is similar to the Major Field Test, though of higher difficulty (see Stoloff & Feeney, 2002). There is a meta-analytic indication that the GRE-P has validity for the prediction of success in master level and doctorate programs (Kuncel, Wee, Serafin, & Hezlett, 2010). Both tests were constructed to assess knowledge of concepts in psychology comprehensively what leads to relatively long tests. The completion of the Major Field Test for Psychology, for example, takes around two hours. The aim of the effort presented here is to create a short test for basic knowledge in psychology which can be used as an indicator of basic knowledge in psychology. In order to make the selection of test content as comprehensible as possible, we referred to textbook analyses trying to identify core concepts in psychology. A short overview of these studies is provided in the following section.

Core Concepts in Psychology

Differences and similarities between introductory psychology textbooks have been the topic of interest of a variety of research efforts. Researchers found that textbooks differ greatly with regard to the number of pages, while the number of chapters and the chapter organization is fairly similar (Griggs & Jackson, 2013; Griggs & Marek, 2001). To create a list of core concepts which are dealt with in nearly all textbooks, several authors compared the boldfaced items in the glossaries of several textbooks. For example, Zechmeister and Zechmeister (2000) examined the boldfaced items in the glossaries, or indexes of 10 introductory textbooks published between 1994 and 1997 and found that the number of terms ranged from 512 to 1,056. The researchers compiled a list of the concepts found in all books. After collapsing similar terms, and terms with the same meaning, 2,505 concepts remained. The authors identified 197 concepts that were included in 80 percent or more of the glossaries.

Similar results were found by Proctor and Williams (2006) who identified 4,902 concepts when searching through the glossaries of 33 textbooks. The researchers found that 428 of these concepts were mentioned in half of the glossaries, while only 33 concepts were dealt with in all glossaries. Griggs, Bujak-Johnson, and Proctor (2004) analyzed 44 textbooks and identified 155 concepts that were dealt with in 80 percent or more of the glossaries. In a different study (Nairn, Ellard, Scialfa, & Miller, 2003), six textbooks were examined. The authors identified 377 concepts that were dealt with in at least five of the six textbooks. When concentrating on those concepts dealt with in all books, the number dropped to 197. Furthermore, the authors compared their list of concepts to the lists compiled by three other studies, resulting in a list of 58 concepts common to all studies. The list contains broadly defined and widely known concepts such as intelligence, classical conditioning, or perception.

In sum, the results show that even though psychology is a broadly diversified field, it is possible to determine a set of concepts that seem to be part of nearly every psychology textbook and can be considered to be a kind of basic knowledge in psychology. This is in line with the guidelines for the undergraduate psychology major provided by the American Psychological Association (2013) arguing that there is a set of basic psychological terminology and concepts that should be mastered by every student.

Development of the Test

As a first step, a list of core concepts was created based on three studies (Griggs et al., 2004; Proctor & Williams, 2006; Zechmeister & Zechmeister, 2000). Only those concepts that were included in at least 50 percent of the examined textbooks were taken into account. These concepts were assigned to one of the three content areas: Natural science oriented, social science oriented and research methods. This classification of test content was adopted from Educational Testing Service (2014b). For the natural science oriented field, ten items referring to the subfields learning, cognition, perception and sensation, as well as biopsychology/neuroscience were created. For the social science oriented field, ten items referring to the subfields clinical/abnormal, developmental, personality, as well as social were created. For the research methods field, six items referring to the subfields research and methods, test theory, as well as statistics were created. In sum, the initial test consisted of 26 items. The items were created to assess individuals' appropriate use of technical terms as well as conceptual understanding. As the optimum item format should be chosen depending on context and purpose (see Martinez, 1999), different formats of items are used. To test conceptual understanding, many items employ Multiple True-False Format (MTF, see Haladyna & Rodriguez, 2013, pp. 72–73). Test takers are confronted with several statements and required to judge separately for each statement whether it is correct. This item format seems to be a viable alternative to conventional multiple choice items with even higher reliability estimates and better time efficiency (Frisbie, 1992; Haladyna & Rodriguez, 2013, p. 73). Other item types used to examine conceptual understanding require the test takers to match terms (i.e. to match examples to concepts), or require the test takers to match a figure to a concept (e.g., “indicate which of the four figures depicts the Yerkes-Dodson-Law”). Items which aim to assess individuals' appropriate use of technical terms predominantly employ a fill-in-the-blank format. This item format is used to examine whether individuals are able to actively recall technical terms in context. The majority of the items cover one certain concept, whereas some items cover several closely related concepts (e.g., an MTF-item to assess knowledge about color perception includes the concepts “Cones” and “Rods”). The maximum score for each item was one point. Partial credit was given if an item was answered partially correct (e.g., some of the true/false-questions constituting an MTF-item were answered falsely).

If one item was not completed at all, it was considered a wrong answer. Items were presented in German; an English translation of four sample items can be found in the appendix. The complete test can be obtained from the authors.

Pilot Study

The purpose of the pilot study was to explore the factor structure of the test and to gain initial experience in the application of the test. Another aim was to examine whether the test would be able to differentiate between psychology students with varying study progress (in terms of year of study).

Methods

The test was handed to a sample of $N = 64$ German psychology students at the University of Trier, Germany, including $n = 22$ freshmen, $n = 21$ advanced undergraduate students as well as master level students, and $n = 21$ PhD students (at German universities, PhD students usually have a master's or an equivalent degree). All participants completed the test individually in a paper-and-pencil format. The mean age of the subjects was $M = 24.67$ ($SD = 3.71$), 78% of them were female.

Results

As the first aim of the pilot study was to explore the factor structure of the test, a principal component analysis with the 26 items was conducted. The Kaiser-Meyer-Olkin-value was .67, suggesting the item set to be suitable for this kind of analysis.

Inspection of the anti-image correlation matrix diagonals showed that these values were below the conventional cutoff value of .50 for five of the items. Four of these items were removed; one item was retained as the value was very close to the cutoff (.48) and because it was one of two items covering the field of biopsychology/neuroscience. The principal component analysis was repeated with the remaining 21 items. The scree plot unambiguously indicated a point of inflexion after the first factor, suggesting a one-factor solution even though seven factors with eigenvalues exceeding 1.0 were extracted (list of eigenvalues: 5.89, 1.83, 1.68, 1.50, 1.19, 1.13, 1.08). This factor explained 28.0% of the variance. Additional support for the assumption of a one-factor-solution came from the inspection of the factor loadings which all exceeded .30 except for the two items covering the field of biopsychology/neuroscience. Therefore, the 21 items were combined into one scale by computing the average of the item scores. Consequently, the total test score can range from 0 to 1. Reliability analyses indicated that the 21-item scale showed an internal consistency of $\alpha = .86$ and split-half-reliability of $r_{tt} = .86$ (Spearman-Brown correction). In this paper, split-half-reliability was computed using the package psych of statistical software R which creates 10000 splits and averages the reliability coefficients obtained. Scores for the three groups defined in the methods section can be found in Table 1. One-way ANOVA with repeated contrasts indicated an overall difference ($F[2, 61] = 53.39, p < .01, \eta^2 = .64$), as well as significant differences between the freshman and the advanced students ($d = -.25, SE = .04, p < .01$), and between the advanced and the PhD students ($d = -.11, SE = .04, p < .01$).

-----insert table 1 here-----

Validation studies

To use the test with larger sample sizes, the 21 items were converted into an online-format that could be completed via online survey software.

The purpose of the first validation study was to examine whether the factor structure could be replicated with a larger sample that allowed for the use of confirmatory factor analysis technique (CFA). Another aim of the study was to determine whether the differences between students with differing study progress could also be replicated with a larger sample. Validation study 2 was conducted to establish a correlation with academic achievement in psychology.

Methods

Validation study 1

The sample of the first validation study consisted of $N = 184$ psychology students from the University of Trier, Germany who completed the test at home together with other measures via online survey software. Data from 2 participants had to be excluded due to technical problems with the survey software; data from 9 participants were excluded as these participants had also participated in the pilot study. The final sample consisted of $N = 173$ psychology students, among them $n = 35$ freshmen, $n = 106$ undergraduate students in their second year or later, and $n = 32$ students enrolled in a master's program or equivalent. Mean age was $M = 22.68$ ($SD = 2.69$); 147 students (around 85%) were female.

Validation study 2

The sample of the second validation study consisted of $N = 114$ second year undergraduate psychology students from the University of Trier, Germany, who took part in a longitudinal study. Data was collected in group settings with around 20 participants in the computer lab of the university. Participants were additionally asked to submit an academic transcript. As two participants refused to submit an academic transcript, the final sample consisted of $N = 112$ students. Mean age was $M = 20.25$ ($SD = 2.20$); 91 of the participants (around 81%) were female.

Results

Validation study 1

A confirmatory factor analysis was computed to examine whether all manifest item variables can be explained by a single latent variable. To account for non-normal distributions of several items, the robust maximum likelihood estimator (MLR) implemented in package lavaan of the statistical software R was used. Overall, fit indices support this model ($\chi^2 [189] = 273.26$; CFI = .88; RMSEA = 0.051; SRMR = 0.060). The value for CFI is slightly below .90 which is conventionally regarded as an acceptable value (McDonald & Ho, 2002); however, both, RMSEA and SRMR indicate good model fit (Hu & Bentler, 1999). Reliability analyses showed a good internal consistency of $\alpha = .84$ and split-half-reliability of $r_{tt'} = .83$ for the total sample. To examine group differences, the sample was split into the three groups defined in the methods section. When examining internal consistency for these subgroups, acceptable values were found ($\alpha = .76$ for the freshmen group, $\alpha = .72$ for the undergraduate students in their second year or later, and $\alpha = .78$ for the master level students). Group means are reported in Table 2. Analysis of variance indicated overall group differences ($F[2,170] = 64.14$, $p < .01$, $\eta^2 = .43$). Repeated contrasts indicated a significant difference between freshmen and advanced undergraduate

students ($d = -.29$, $SE = .03$, $p < .01$), but no significant difference between advanced undergraduate students and master level students ($d = -.03$, $SE = .03$, ns).

-----insert table 2 here-----

Validation study 2

In a first step, reliability analyses were conducted. Internal consistency ($\alpha = .52$) and split-half-reliability ($r_{tt} = .54$) were markedly lower than in validation study 1. To identify possible reasons for this result, descriptive test and item statistics were inspected. On average, the participants reached a score of $M = 0.65$ which is comparable to the group of advanced undergraduate students from validation study 1 while the variability of scores is slightly lower. As can be seen in Table 2, the standard deviation for the total sample of validation study 1 was $SD = 0.18$ while the same parameter was $SD = 0.10$ in validation study 2. Thus, sample homogeneity might be one reason for lower reliability of the test in validation study 2. In addition, the variability of item difficulties was considerably higher in validation study 2 which leads, in turn, to a decrease of inter-item-correlations, lower homogeneity of the test, and lower item-total correlations: Mean scores varied between $M = .02$ and $M = 0.88$, and nine of the 21 items had mean scores of $M > .80$, that is, they were quite easy for the participants. Contrastingly, in validation study 1 most item means were in the medium range ($M = .17$ to $M = .78$), and only one item had a mean score of $M > .80$.

Because all students were in their second year, it was not possible to examine group differences as in validation study 1, or the pilot study.

The main objective of validation study 2 was to test for correlations with academic achievement in psychology. Preliminary college GPA was determined by computing the mean of all course grades obtained until the time of the study. The average grade was $M = 2.06$ ($SD = 0.58$). According to the German grading system, grades may range from 1 (very good) to 4 (satisfactory performance). As smaller grade numbers indicate higher performance, the highly significant correlation between preliminary college GPA and test score was negative ($r = -.41$, $p < .01$).

5. Discussion

The results indicate that the test can be used as a reliable and valid instrument to assess basic psychology knowledge among populations with varying knowledge level. Even though the test items were selected based on a comparatively small sample, the one-factorial structure of the test could be confirmed by the results of a confirmatory factor analysis based on data from a larger sample. Internal consistency and split half-reliability reached satisfactory levels in the pilot study and the first validation study, even though the items cover material from a broad range of fields of psychology. Reliability was substantially lower in validation study 2, a result which was disappointing and is not easily interpreted. Apart from sample bias as a possible reason, it might be argued that the sample was more homogenous as it consisted exclusively of second year undergraduate students what slightly reduced the variance of test scores. As mentioned above, an increased number of items with extreme item difficulties might have reduced the internal consistency. A speculative but plausible interpretation of this finding is that most of these easy items referred to concepts which were part of the content of exams which took place shortly before the data of the validation study 2 were collected.

Differences between groups with differing study progress can be interpreted as indication of validity for the test as the items cover material that is learned during the entire undergraduate curriculum. It should be noted that there are no introductory courses in psychology at the University of Trier (similar to the situation at most other German universities), so many first year students have never heard of terms which are part of the second or third year curriculum. Furthermore, advanced students had more opportunities to accumulate the relevant knowledge. These findings are in line with the findings by Pinter et al. (2014) who found significant differences between graduating students, and freshmen on the Major Field Test for psychology and by Dolinsky and Kelley (2010) who found a correlation between the credits earned in psychology and performance on the Major Field Test. Concerning indications of validity for the test, however, the more important finding is certainly the substantial correlation with the course grades established in validation study 2. It indicates that the test shows moderate convergent validity with academic achievement in psychology.

An issue which is generally pertinent when using fixed-choice-items is the distortion of results by guessing, especially when there are only two response options like with MTF-items. The test, however, also contains several items using the fill-in-the-blank format that do not allow guessing. If scores on the MTF-Items had been substantially distorted by guessing, intercorrelations among the items (as summarized by split-half-reliability and internal consistency) would not be of the magnitude reported above. We therefore expect guessing not to be a major issue with our test.

It should be noted that, despite good reliability, this test should not be used for individual high stakes decisions as the number of items is too low and reliability scores do not justify an application in individual diagnostics. Additionally, due to the small number of items, subscales related to different fields of psychology should not be computed and used for individual assessment. Admittedly, the creation of such a test was not our intention; instead we tried to create a short test that offers a snapshot of knowledge of psychology concepts. The main purpose of our efforts was to create a test that can be used as indicator of psychology knowledge in research studies where time available for data collection is limited and longer tests, like the Major Field Test (Educational Testing Service, 2014a), cannot be used.

It can also be asked whether the test assesses pure knowledge of facts, or understanding of psychology concepts. Several authors argue that fixed-choice-tests capture mainly the effects of rote learning, instead of understanding, or thinking skills (e.g., Tung, 2010). Other authors argue that it is possible to create fixed-choice-tests that tap deeper levels of thinking (American Psychological Association, 2009). We believe that we have achieved this by using differing item formats; however, there is no evidence. As probably every test, this test certainly also measures verbal skills, as understanding the questions certainly requires an amount of verbal skills; for example, Dolinsky and Kelley (2010) showed that performance on the Major Field Test could be predicted by the verbal SAT score.

Another criticism of standardized tests is that test results can be distorted by student course selection as the tests might include materials that have not been learned by every student (American Psychological Association, 2009). This issue should certainly be kept in mind when using a standardized test. In case of the present test, we do not consider this problematic, as the test items cover basic knowledge that is part of nearly every curriculum.

Finally, several limitations should be mentioned. First, it can be questioned whether it is appropriate to use a list of core concepts based on US textbooks to create a test for German

students. However, many US textbooks have either been translated into German or are used in their original version for teaching purposes in Germany. Second, it should be mentioned that lists of core concepts do not indicate in what detail each concept is dealt with in textbooks. Therefore, some items might be very hard to answer for students. Third, it can be questioned whether the findings from the three studies are comparable, as test format (paper and pencil vs online) and data collection setting (at home vs in a supervised setting) differ. However, the results indicate that test scores are not affected by these differences: The relationship with study progress and satisfactory reliability estimates found in the pilot study could be replicated in validation study 1. Finally, considering the comparatively large samples size in the two validation studies, it can be questioned whether Rasch modeling might have been an option. The main advantage of Rasch modeling (the interchangeability of equally difficult items) though, would not be applicable in the present case as the test contains only 21 items.

In addition to addressing these limitations, future research efforts might also assess participants' intelligence to examine the relationship of this test with general intelligence, and conduct longitudinal studies to examine retest reliability and predictive validity of this test with regard to the grade awarded for the final thesis, for example.

Conclusion

To sum up, a short test for the development of basic knowledge in psychology has been developed. Our findings demonstrate satisfactory levels of reliability estimates and indicate that the test can be used in paper-and-pencil format, or in an online format. The group differences and the relationship with academic achievement point to the validity of the instrument. The test contains items of varying difficulty and can obviously be used with populations with great variance in levels of psychology knowledge. This makes the test a useful instrument when an economically employable test for basic knowledge in psychology is needed.

References

- American Psychological Association. (2009). *The assessment cyberguide for learning goals and outcomes*. Retrieved from <http://www.apa.org/ed/governance/bea/assessment-cyberguide-v2.pdf>
- American Psychological Association. (2013). *APA guidelines for the undergraduate psychology major: Version 2.0*. Retrieved from <http://www.apa.org/ed/precollege/undergrad/index.aspx>
- Dolinsky, B., & Kelley, J. M. (2010). For better or for worse: Using an objective program assessment measure to enhance an undergraduate psychology program. *Teaching of Psychology, 37*, 252–256. doi:10.1080/00986283.2010.510978
- Educational Testing Service. (2014a). *ETS Major Field Test for Psychology*. Retrieved from <http://www.ets.org/mft/about/content/psychology>
- Educational Testing Service. (2014b). *GRE Subject Test Psychology*. Retrieved from <http://www.ets.org/gre/subject/about/content/psychology>
- Frisbie, D. A. (1992). The Multiple True-False item format: A status review. *Educational Measurement: Issues and Practice, 11*(4), 21–26. doi:10.1111/j.1745-3992.1992.tb00259.x
- Gallagher, S. P., & Cook, S. P. (2013). The validity of the Major Field Test in Psychology as a programme assessment tool. *Psychology Teaching Review, 19*, 59–72.
- Griggs, R. A., Bujak-Johnson, A., & Proctor, D. L. (2004). Using common core vocabulary in text selection and teaching the introductory course. *Teaching of Psychology, 31*, 265–269. doi:10.1207/s15328023top3104_8
- Griggs, R. A., & Jackson, S. L. (2013). Introductory psychology textbooks: An objective analysis update. *Teaching of Psychology, 40*, 163–168. doi:10.1177/0098628313487455
- Griggs, R. A., & Marek, P. (2001). Similarity of introductory psychology textbooks: Reality or illusion? *Teaching of Psychology, 28*, 254–256. doi:10.1207/S15328023TOP2804_03
- Haladyna, T. M., & Rodriguez, M. (2013). *Developing and validating test items*. New York, NY: Routledge.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. doi:10.1080/10705519909540118
- Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement, 70*, 340–352. doi:10.1177/0013164409344508
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist, 34*, 207–218. doi:10.1207/s15326985ep3404_2
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64–82. doi:10.1037//1082-989X.7.1.64
- Nairn, S. L., Ellard, J. H., Scialfa, C. T., & Miller, C. D. (2003). At the core of introductory psychology: A content analysis. *Canadian Psychology/Psychologie Canadienne, 44*, 93–99. doi:10.1037/h0086930
- Pinter, B., Matchock, R. L., Charles, E. P., & Balch, W. R. (2014). A cross-sectional evaluation of student achievement using standardized and performance-based tests. *Teaching of Psychology, 41*, 20–27. doi:10.1177/0098628313514174

- Proctor, D. L., & Williams, M. E. (2006). *Frequently cited concepts in current introduction to Psychology textbooks*. Retrieved from <http://www.plainlocal.org/userfiles/885/Classes/20186/List%20of%20commonly%20used%20terms%20in%20psych%20books-0.pdf>
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory. *Perspectives on Psychological Science, 1*, 181–210. doi:10.1111/j.1745-6916.2006.00012.x
- Stoloff, M. L., & Feeney, K. J. (2002). The major field test as an assessment tool for an undergraduate psychology program. *Teaching of Psychology, 29*, 92–98. doi:10.1207/S15328023TOP2902_01
- Tung, R. (2010). *Including performance assessments in accountability systems: A review of scale-up efforts*. Boston, MA: Center for Collaborative Education.
- Zechmeister, J. S., & Zechmeister, E. B. (2000). Introductory textbooks and psychology's core concepts. *Teaching of Psychology, 27*, 6–11. doi:10.1207/S15328023TOP2701_1

Table 1*Group differences in the pilot study*

group	<i>n</i>	<i>M</i>	<i>SD</i>
first year students	22	0.44	0.13
advanced B.Sc. and M.Sc. students	21	0.69	0.11
Ph.D. students	21	0.79	0.11
overall mean	64	0.64	0.19

Table 2*Group differences in validation study 1*

group	<i>n</i>	<i>M</i>	<i>SD</i>
first year students	35	0.35	0.14
second year students	106	0.64	0.13
M.Sc. or equivalent	32	0.67	0.15
overall mean	173	0.59	0.18

Appendix sample items from the psychology knowledge test¹

1. Indicate whether the following statements are true or false

Color perception...

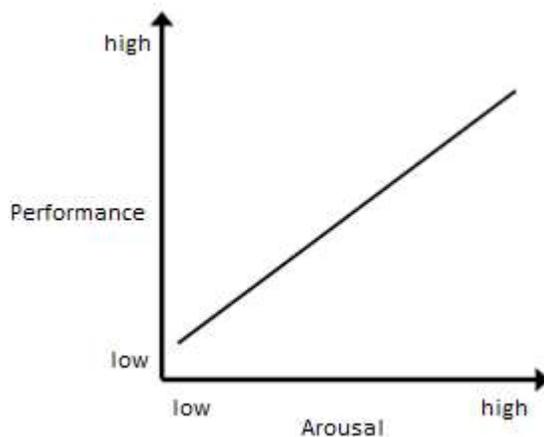
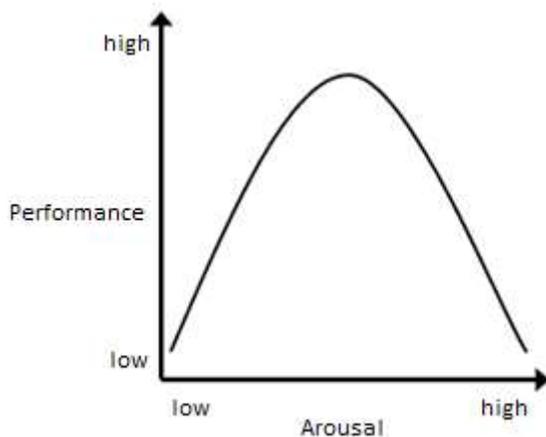
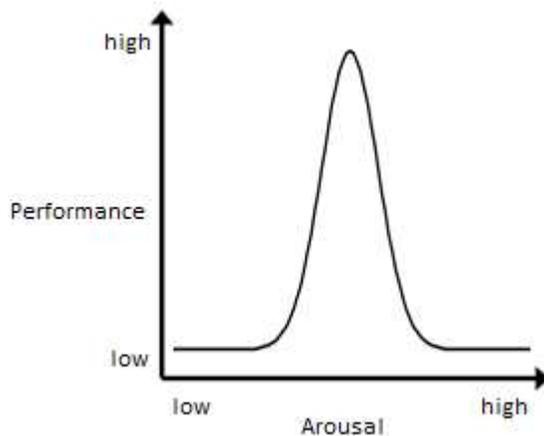
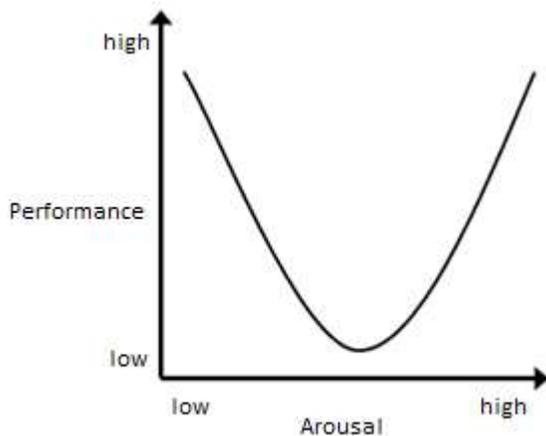
	true	false
- takes place in photoreceptors	x	o
- occurs in cones	x	o
- occurs in rods	o	x
- is determined by the wavelength of light	x	o

[x = correct answer; o = wrong answer]

2. Complete the following sentence:

The motivation to do something as a goal in its own right is called _____ [intrinsic] motivation.

3. The Yerkes–Dodson law refers to the relationship between cognitive performance and physiological activation. Indicate which of the following four figures depicts this relationship correctly.



¹ The complete test can be obtained from mayer@zpid.de

[Correct answer: left bottom]

4. Match the following terms to the correct modes of processing.

Mode of processing	Terms
Top-down	hearing (bu) concepts (td) previous experience (td) vision (bu)
Bottom-up	knowledge (td) expectation (td)

[Correct responses (bold): bu = bottom-up; td = top-down]