

Wright State University

CORE Scholar

International Symposium on Aviation
Psychology - 2009

International Symposium on Aviation
Psychology

2009

How Effective Is Item Bank Testing of Pilot Training Applicants in Reducing Test Preparation Effects?

Oliver Zierke

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2009



Part of the [Other Psychiatry and Psychology Commons](#)

Repository Citation

Zierke, O. (2009). How Effective Is Item Bank Testing of Pilot Training Applicants in Reducing Test Preparation Effects?. *2009 International Symposium on Aviation Psychology*, 57-62.
https://corescholar.libraries.wright.edu/isap_2009/105

This Article is brought to you for free and open access by the International Symposium on Aviation Psychology at CORE Scholar. It has been accepted for inclusion in International Symposium on Aviation Psychology - 2009 by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

HOW EFFECTIVE IS ITEM BANK TESTING OF PILOT TRAINING APPLICANTS IN REDUCING TEST PREPARATION EFFECTS?

Oliver Zierke
German Aerospace Center (DLR)
Aviation and Space Psychology
Hamburg, Germany

In the selection of aviation personnel, special test preparation has become an emerging problem. Specific test preparation aims at raising the probability to master a certain test rather than developing the underlying ability. Knowledge tests are particularly susceptible to the problem of test preparation. One strategy to counter this problem is the use of comprehensive item banks for testing in knowledge domains. In 2005 over 770 student pilot applicants participated in an evaluation study of two item bank tests, an English language test and a test of physical knowledge. A conventional test form as well as an item bank test form were given to each subject. Consequently, both test forms were compared in a repeated measures design. The test preparation effects, correlations with school grades, and prognostic validity of both tests were analyzed. It is shown that item bank testing reduces test preparation effects and enhances construct validity.

Parallel to the rapid changes in the aviation business, a new challenge in the selection of student pilots must be realised: The problem of test preparation offered through new media, such as the internet, or by commercial training institutes. This kind of specific test preparation aims at raising the probability to master a certain test rather than developing the underlying ability. For a pilot training applicant, the successful accomplishment of a selection procedure can result in sponsored flight training, financed by a few larger commercial airlines. Compared to a private pilot training, such sponsorships can provide a suitable applicant with several ten thousand Euros worth of training. Testing in aviation is therefore referred to as “high stake testing”. Thus, it is quite understandable that applicants are willing to try everything to prepare optimally, and a test preparation market has evolved for satisfying this need. In Germany at least four commercial institutes, one commercial CD with training material, and two internet chat rooms exist exclusively for the preparation for the DLR (German Aerospace Center) test. For the applicant, as well as for the preparation institute, it matters little whether the student actually improves his aptitude or general knowledge or whether he simply improved his ability to solve one specific test. The latter would be the case if an applicant has access to the questions of a test, e.g. a technical comprehension test, prior to taking it. He could possibly memorize these very items and their correct solutions without any in-depth understanding of the subject.

Test preparation effects are defined as achieving higher scores without real knowledge of the underlying domain. These effects lead to an overestimation of the ability of a dishonestly prepared candidate. This example demonstrates the threat of test preparation for selection in aviation business, because test fairness and test validity can be compromised. This can lead to incorrect selection decisions, which in the long run may have an impact on aviation safety. For this reason countermeasures are necessary.

Countermeasures

The problem of specific test preparation concerning aptitude tests is answered by constructing new tests regularly and by offering detailed pre-information and own training material to the applicants (Huelmann & Oubaid, 2004). Countermeasures concerning knowledge tests differ from those for aptitude tests. Knowledge tests are frequently used for licensing purposes (Impara, 1995) or for measuring basic requirements for an apprenticeship. Therefore, knowledge tests play a prominent role in aviation psychology. The problem of test preparation is of particular importance for knowledge tests, because it is not difficult for applicants to publish via internet memorized items from the test after completing the examination and to provide future applicants with preparation material. A method to counter the preparation problem for knowledge tests is the use of comprehensive item banks instead of fixed tests. Using item banks lowers the predictability of items for test takers and thus may encourage them to prepare for the whole subject of the test rather than merely for the known individual items.

The approach of DLR

At the German Aerospace Center, item banks were installed for the knowledge domains of physics, mechanical comprehension, mathematics, and English language. For every individual test form, items are randomly drawn from the item bank, while maintaining a balance of item difficulty, test standard deviation and reliability for all forms (Figure 1). This procedure is based on a method developed by Gibson and Weiner (1998) and leads to different test forms for each applicant.

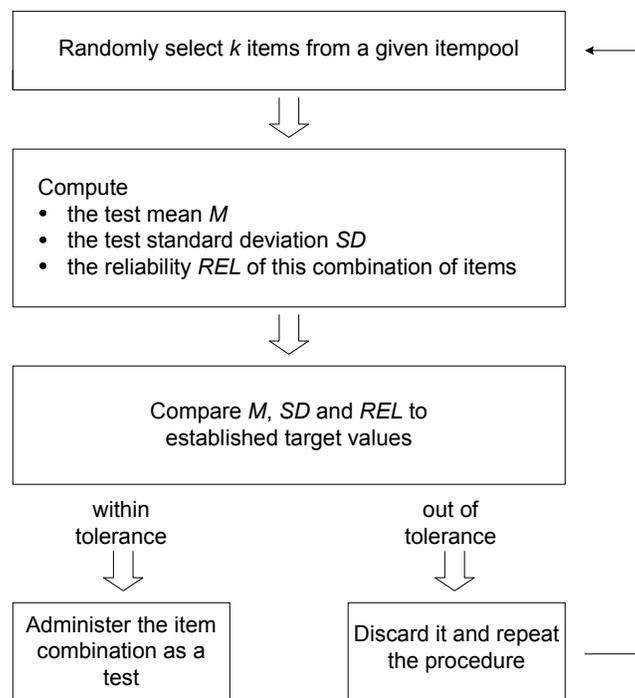


Figure 1. The procedure of test assembly

Method

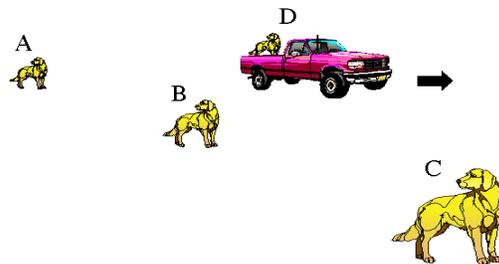
An evaluation study of item bank testing as a means of reducing test preparation effects was conducted. In 2005 over 770 student pilot applicants participated in this study. The English item bank was composed of four parallel tests which were active during the past in the DLR pilot selection. This assembly resulted in an item bank comprising 204 items. A single test drawn out of this item bank consisted of 60 items and had an internal consistency of Cronbach's $\alpha = .90$. The item bank of physical knowledge consisted of 104 completely new items. A resulting single test comprised 40 items and had an internal consistency of Cronbach's $\alpha = .78$. Item examples are shown in Figure 2. A conventional test form as well as an item bank test form were given to each subject. Consequently, both test forms were compared in a repeated measures design.

English Test

To participate the Olympics must be a real thrill.

- 1) on
- 2) in
- 3) by
- 4) at

Physical Knowledge Test



“Which dog hears the sound of the pickup truck in the highest frequency?”

Figure 2. Item examples for English and Physical Knowledge Test

Subjects were requested in a questionnaire to disclose any commercial preparation. In this study a test preparation effect is calculated as the mean difference in test scores between the two groups of candidates, one which was commercially prepared and another which was not. The following hypotheses have been addressed:

Hypotheses

1. Test preparation effects are smaller for item bank tests when compared with those of conventional tests.
2. Item bank tests show larger correlations with school grades than the conventional tests do.
3. Item bank tests show higher prognostic validity than conventional tests.

Results

Test preparation

For the English test 34 of 451 applicants disclosed they have attended a commercial preparation course. This makes a preparation rate of 7.5%. For the Physical Knowledge Test only 16 of 314 applicants disclosed a commercial preparation course. This means a preparation rate of 5.1%. The English and the Physical Knowledge Test have been administered at different times, thus the difference in the preparation rates is explainable.

Hypothesis 1

The first two hypotheses were confirmed completely. Item bank testing reduces test preparation effects for both the English and the Physical Knowledge Test (see Figure 3 and 4). For both tests the ANOVA interaction effect became highly significant with $F(1, 449) = 40.0$ for English, and $F(1, 312) = 46.4$ for Physics. This resulted in a medium effect for the English test ($f = 0.30$) and a large effect for the Physical Knowledge test ($f = 0.39$).

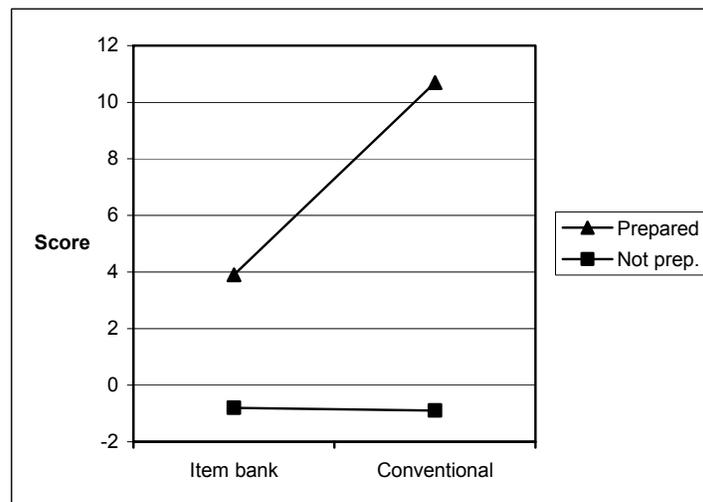


Figure 3. Test preparation effects for English Test

For not specially prepared applicants there was no difference, whether they got an item bank test or a conventional test form. In both tests they reached nearly the same result. In contrast, specially prepared applicants achieved much higher scores in the conventional tests, presumably because they already knew some items. Therefore, item bank tests yield more realistic measurements of aptitudes for prepared applicants in particular.

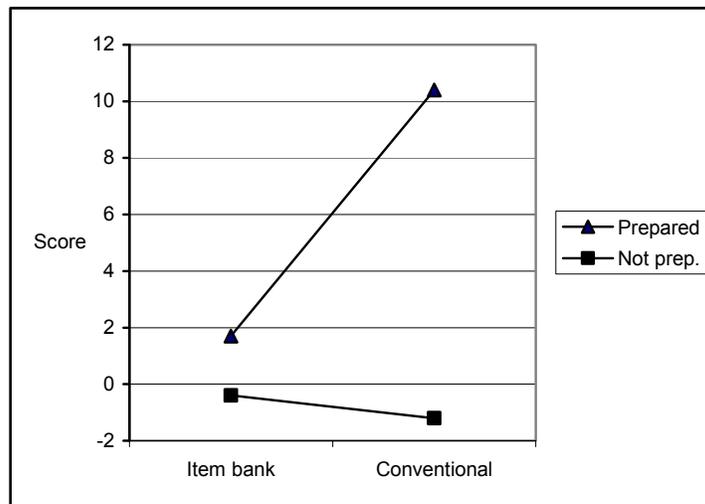


Figure 4. Test preparation effects for Physical Knowledge Test

Hypothesis 2

Item bank testing enhances construct validity in form of correlations with school grades. The respective correlation for the item bank test is significantly higher than for the conventional test ($r = .51$ vs. $r = .43$, $N = 379$, $p < .01$ for the English test and $r = .39$ vs. $r = .25$, $N = 266$, $p = .01$ for the Physical Knowledge test). That means that item bank tests measure more true variance than conventional tests.

Hypothesis 3

For a definite confirmation concerning the third hypothesis, the data could not be interpreted clearly because too few applicants were recommended for pilot training to calculate stable correlations.

Discussion

Item bank testing turned out to reduce test preparation effects in the selection of pilot training applicants. This is an important result because commercial test preparation is a challenge to test fairness as well as to test validity. Consequently, the second question was whether reduced test preparation effects will improve test validity. Indeed, item bank testing raised the correlations with school grades as an aspect of construct validity, which means that item bank tests measure more true variance than conventional fixed tests. This result is not surprising. It seems obvious that large item banks reduce the predictability of items for prepared applicants and thus improve the quality of measurement in terms of test fairness and validity. It has never been shown before how effective item banks are in contrast with conventional tests. The item banks reduced test preparation effects, although not to zero. Prepared applicants are still better than not specially prepared ones. Why? The question is whether these differences are true differences, e.g. if prepared candidates really learned and understood more than the unprepared group. If so, they must achieve better results. It seems plausible that candidates who invest more time in their preparation are on average more

motivated and consequently achieve better results. Therefore, we should not aim for tests yielding equal results for prepared and unprepared applicants. Rather, we should ensure that possible differences between both groups represent true differences. With regard to the third hypothesis further research is needed to learn more about the effects of item bank testing on the prognostic validity of knowledge tests.

References

- Gibson, W. & Weiner, J.A. (1998). Generating random parallel test forms using CTT in a computer-based environment. *Journal of Educational Measurement*, 35, 297-310.
- Huelmann, G. & Oubaid, V. (2004). Computer assisted testing (CAT) in aviation psychology. In: K.-M. Goeters (Ed.), *Aviation psychology: practice and research* (pp. 123-134). Aldershot: Ashgate.
- Impara, J.C. (Ed.) (1995). *Licensure testing: purposes, procedures, and practices*. Lincoln: Buros institute of mental measurement.