Wright State University

CORE Scholar

International Symposium on Aviation Psychology - 2021

International Symposium on Aviation Psychology

5-1-2021

Using a Perceptual Speed Test to Predict Flight Training Performance: New Findings

Diane L. Damos

Alvah C. Bittner

Mary Niemczyk

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2021

Part of the Other Psychiatry and Psychology Commons

Repository Citation

Damos, D. L., Bittner, A. C., & Niemczyk, M. (2021). Using a Perceptual Speed Test to Predict Flight Training Performance: New Findings. *34th International Symposium on Aviation Psychology*, 30-35. https://corescholar.libraries.wright.edu/isap_2021/6

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 - 2021 by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

USING A PERCEPTUAL SPEED TEST TO PREDICT FLIGHT TRAINING PERFORMANCE: NEW FINDINGS

Diane L Damos, Ph.D.
Damos Aviation Services, Inc.
Gurnee, IL
Alvah C Bittner, Ph.D., CPE
Bittner and Associates
Kent, WA
Mary Niemczyk, Ph.D.
Arizona State University
Mesa, AZ

Perceptual speed tests have been successfully used for US military pilot selection for 80 years yet are rarely used by other militaries or for civilian pilot selection. This paper describes a study examining the predictive validity of a perceptual speed test, the Tabular Speed Test (TST), for civilian flight training. The TST was administered to university students as they entered a professional pilot curriculum. The number correct (COR) correlated significantly with time to solo (N = 119, r = -.15, p = .05) and with time to private pilot's certificate (N = 51, r = -.34, p = 0.01) but not with graduation/elimination from the pilot curriculum. The number of incorrect responses (WR) did not correlate with any performance measure. Average test-retest reliability from 5 to 17 months was r =.604 for COR and r = .431 for WR, p < .001 for both.

The first use of a perceptual speed test for pilot selection occurred in December 1942 when the Dial Reading and Table Reading Tests were included in the US Army Air Forces aircrew classification battery (Guilford & Lacey, 1947). The Dial Reading Test was eventually dropped, leaving the Table Reading Test as the only perceptual speed test. This test is still in the US Air Force pilot selection battery (Carretta & Ree, 1995; Johnson, Barron, Carretta, & Rose, 2017). Despite the well-documented usefulness of a perceptual speed test for over 75 years, only the Norwegian Air Force includes a perceptual speed test in its pilot selection battery (Martinussen & Torjussen, 2004).

Like their military counterparts, civilian pilot selection batteries rarely include a perceptual speed test. One exception to this is the German Aerospace Center (DLR) selection battery that was developed for civilian ab initio pilot selection (Goeters, Hoermann, & Maschke, 1989). Another exception is the selection battery for entrance into the professional flying curriculum of the Artic University of Norway (M. Martinussen, personal communication, February 22, 2021).

In 2008, Mount, Oh, and Burns (2008) demonstrated that, for perceptual speed tests, the number correct (COR) and the number wrong (WR) assess different attributes. They suggested that, whereas COR assesses speed of processing, WR reflects an inability to focus on the task at hand and a tendency towards "distractibility, carelessness, recklessness, or apathy on the job (p.118)". They suggested that WR may predict problems with rules compliance, which would be manifested in "accidents, safety violations, tardiness, and use of alcohol or drugs on the job (p. 118)." If this were true, then pilot selection batteries should include a perceptual speed test, and both WR and COR should be included in any selection decision.

This study had two major goals. The first was to confirm the Mount et al. results that COR and WR measure different attributes. We sought to confirm Mount et al.'s results in three ways. First, we obtained test-retest data at three different intervals. If COR and WR measure stable attributes, they should demonstrate significant reliabilities with little decrease over time. Second, assuming that the COR and WR measure different attributes, their correlation should be low. Third, again assuming that COR and WR assess different attributes, they should correlate with different behaviors.

The second goal was to determine the predictive validity of COR and WR. US civilian ab initio training has three major milestones: time to solo, time to the private pilot certificate (PPC), and program completion. This study evaluated the predictive validity of COR and WR for all three milestones.

Approach

A perceptual speed test, the Tabular Speed Test (TST), was administered each semester beginning with the fall semester of 2005 through the spring semester of 2008 to all students enrolled in the aviation program of a large western university. The college offered three majors: professional flight (pilot), air traffic control, and aviation management. To obtain retest data, the test also was administered once during this 3.5-year period in an advanced course typically taken by third- and fourth-year professional flight and aviation management students and once during a departmental safety meeting that was mandatory for all flight students.

Mekhail, Niemczyk, Ulrich, and Karp (2010) obtained the time to solo and time to PPC data presented in this study. Our data analysis differs in that we limited our analyses to those students with known outcomes, i.e., those definitely identified as either having dropped out of the curriculum or graduated from the curriculum. We also analyzed WR, which they did not.

Mekhail et al. (2010) observed that the completion times reported by some students appeared to be estimates. During the time of our study, the department kept a student's flight records for only two years after the student either graduated or left the program. Thus, it was not possible to obtain a more accurate measure of time to solo or time to PPC for many of the students. Additionally, not all data were available for all students, i.e. a student might report time to PPC but not time to solo. These estimates and the unknown cause for dropping out of the flight curriculum resulted in greater random error than in a more controlled study. We adopted a one-tailed $p < .10$ level of statistical significance rather than the more common, two-tailed $p < .05$ (Wickens, 2015) for testing the three milestones because of the greater level of random error and because prior studies provided directional expectations concerning the relation between the three milestones and the two dependent measures.

Methods

Subjects

The TST was administered to all students taking the two classes described earlier and attending the departmental flight safety meeting. Participation was voluntary; students who did not want to participate did not complete the test form. Students were given no financial incentives to participate nor did they receive any additional course credit.

Tabular Speed Test

The TST is a paper-and-pencil, 50 question, multiple-choice test. The test has a 9-minute time limit. The TST has two parallel versions (equivalent means, standard deviations, and distributions), both of which were used.

Administration

The TST forms were distributed at the beginning of class or at the beginning of the safety meeting. Approximately equal numbers of each version were distributed during each testing session. The test administrator started and stopped the testing period but provided no information on the time remaining during the test. None of the test administration classrooms had clocks. The test administrator did not control which version of the TST a testee received for the retest. All test administrations were conducted by the third author.

Results

Although all students enrolled in the aviation program took the introductory class, only data from students enrolled in the professional flight curriculum were analyzed. Both COR, time to solo, and time to PPC had normal distributions. WR, however, was skewed and leptokurtic. A *ln* (WR+1) transformation resulted in acceptable distributions for all of the analyses. Unless otherwise indicated, all analyses were conducted using Excel 365.

COR Versus WR

Test-retest. Because of differences in student schedules, the retest interval varied. Three intervals had sufficient data to allow reliability calculations: 5 months, 12 months, and 17 months. Data for the 5-month interval was obtained from students who took the retest the following semester during the departmental safety meeting or who repeated the introductory class because they dropped the class or failed it. Data for the 12- and 17-month interval came from students who took the retest during the safety meeting or during the advanced course. The results shown in Table 1 are Pearson Correlations.

Table 1.

Test-retest reliabilities as a function of testing interval

Testing interval in months	N	COR	Ln(WR+1)
5	41	.70****	.29
12	24	.45*	.53**
17	23	.64***	.45*

** **p<.0001, *** p< .001. ** p< .01, * p< .05

If COR and *ln* (WR+1) test-retest reliabilities represent attributes that are stable across time, then their respective reliabilities should be essentially constant over time. If they assess attributes that are affected by experience, then their reliabilities should decrease over time. Table 1 does not show either pattern clearly, though there is a slight suggestion of a break after 5 months. Consequently, we sequentially tested (Lenhard & Lenhard, 2014) both the Olkin-Pratt averaged 12-17 month COR correlation against the 5 month correlation after we tested the 12 versus 17 month correlation; neither contrast was significant ($p > 0.10$ for both). This process was repeated for the *ln* (WR +1) correlations, and again neither contrast was significant ($p > 0.10$). Finally, we averaged the respective correlations using the Olkin and Pratt (1958) weightings. The averaged reliabilities were .604 for COR (weighted N estimate = 86, $p < 0.00001$) and .431 for *ln* (WR+1) ($p < 0.0001$).

Correlation between COR and WR. To calculate the correlation between COR and WR, first-administration data were used. The correlation was $r = -.19$, $N = 144$, $p = 01$.

Performance Measures

Time to solo. The first major milestone in ab initio flight training is time to solo. We performed directional tests on COR and WR because we wished to determine if time to solo decreased as COR increased (negative relation) and increased as WR increased (see Bittner, Bittner, Winn, & Lundy, 2004 for a discussion of the use of directional testing). The correlation between COR and time to solo was $r = -0.15$, $N = 119$, $p = 0.05$. The corresponding value for *ln* (WR+1) was $r = 0.01$, $N = 119$, $p = 0.46$.

Time to PPC. The second major milestone is time to PPC. Again, we used directional testing to determine if time to PPC decreased with increasing COR (negative relation) and increased with increasing WR. For this milestone, the correlation between time to PPC and COR was $r = -.34$, $N = 51$, $p = .01$. For *ln* (WR+1) the corresponding values were $r = -.06$, $N = 51$, $p = .34$.

Completion of program. The completion status of each student (graduated from the professional flight curriculum or dropped out) was obtained by examining the university's records. We obtained graduation data on 144 students. Of these, 94 completed the professional flight curriculum and 50 did not. We performed two directional (one-tailed) comparisons (Bittner et al., 2004). The first determined if COR for those who completed flight training was significantly greater than for those who dropped out. The second determined if WR was greater for those who dropped out versus those who completed the curriculum. *T*-tests for unequal variances conducted on COR and *ln* (WR +1) were both nonsignificant (t (90) = 0 .98, p = .16 and t (90) = .58, p = .28, respectively).

Discussion

This study had two goals. The first of these was to confirm Mount et al's (2008) finding that COR and WR measure different attributes. We used three different methods to confirm their results. The first was to demonstrate that COR and WR have significant and stable test-retest reliabilities over extended time intervals. Table 1 shows that COR had significant reliabilities up to 17 months after initial testing. WR showed a similar pattern although the reliabilities were generally lower but still statistically significant. The second method concerned the correlation between COR and WR. Mount et al. found a significant correlation of -.46 between these two variables. We used only the first administration data to make our results comparable to those of Mount et al. and found a significant correlation of -.19. The third was to contrast the behaviors predicted by COR versus WR. Mount et al. (2008) argued that COR measures performance, whereas WR measures compliance behaviors. In this study COR correlated significantly with two of the three performance measures, time to solo and time to PPL; WR correlated with none of the three performance measures. Because this study had no measures of compliance behavior, we cannot determine which attributes may be measured by WR. Consequently, we were only partially successful in supporting Mount et al.

The second goal of this paper was to determine the predictive validities of COR and WR for three milestones in ab initio student training: time to solo, time to PPC, and program completion. COR successfully predicted time to solo and time to PPC but failed to predict program completion. The department had no information on why a student dropped out of the professional flying program. Thus, we could not distinguish between those students who were failing, found that they disliked flying, or dropped out for financial reasons. The last administration of the TST occurred during the spring 2008 semester. By this time, the economic recession of 2008 had begun, which may have caused more students than usual to drop out for financial reasons.

In summary, COR predicted both time to solo and time to PPC, which supports the use of perceptual speed tests in civilian, as well as military, pilot selection batteries. The correlation between COR and WR and the constant test-retest reliabilities provide good support for Mount et al's (2008) findings that COR and WR assess different attributes. COR clearly assesses performance, but the attributes assessed by WR must await further research.

References

Bittner, A. C. J., Bittner, R. C., Winn, F. J., & Lundy, N. C. (2004). Theory-based directional testing: Revisiting a primary option for efficient human factors/ergonomics (HF/E) studies. In *Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 2466-2470). Washington, DC: Human Factors and Ergonomics Society.

Carretta, T. R., & Ree, M. J. (1995). Air force officer qualifying test validity for predicting pilot training performance. *Journal of Business and Psychology, 9*(4), 379-388.

Goeters, K.-M., Hoermann, H.-J., & Maschke, P. (1989). The DLR test system ab initio pilot selection. In *Fifth International Symposium on Aviation Psychology* (pp. 663-668). Columbus, OH: Department of Aviation, the Ohio State University.

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Printed classification tests Report No. 5*. Washington, D.C.: U.S. Government Printing Office.

Johnson, J. F., Barron, L. G., Carretta, T. R., & Rose, M. R. (2017). Predictive validity of spatial ability and perceptual speed tests for aviator training. *International Journal of Aerospace Psychology, 27*(3-4), 109-120.

Lenhard, W., & Lenhard, A. (2014). *Hypothesis Tests for Comparing Correlations.* Bibergau, Germany: Psychometrica. doi:DOI: 10.13140/RG.2.1.2954.1367.

Martinussen, M., & Torjussen, T. M. (2004). Initial validation of a computer-based assessment battery for pilot selection in the Norwegian Air Force. *Human Factors and Aerospace Safety, 4*(3), 233-243.

Mekhail, A., Niemczyk, M., Ulrich, J., & Karp, M. (2010). Using the Table Reading Test as an Indicator for Success in Pilot Training. *Collegiate Aviation Review, 28*(1), 101-114.

Mount, M. K., Oh, I., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology, 61*, 113-139.

Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlaton coefficients. *Annals of Mathematical Statistics, 29*, 201-211.

Wickens, C. D. (2015). Statistical errors in aviation psychology: Commonsense statistics in aviation safety research. In *18th International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.