Journal of Aviation/Aerospace Education & Research

Volume 24
Number 2 *JAAER Winter 2015*

Article 2

Winter 2015

Consistency of Aviation Students When Taking the Private Pilot Knowledge Exam

Scott R. Winter
Florida Institute of Technology, scott.winter@mac.com

Stephen Rice
Florida Institute of Technology, srice@fit.edu

Katie M. Reid
Florida Institute of Technology, kreid2013@my.fit.edu

Rian Mehta
Florida Institute of Technology, rianmehta91@gmail.com

Introduction and Background

Assessments are commonly used to help measure student performance (Kubiszyn & Borich, 2010). When it is correctly administered and completed, an assessment helps to distinguish between students that perform satisfactorily and those that do not. Originally, there was an assumption that students' separation in scores was a result of knowledge acquired on the specific topic. Previous studies revealed the potential performance of students on assessment completion when adjusting for consistency (Rice, Geels, Trafimow, & Hackett, 2011; Rice, Trafimow, & Kraemer, 2012). These previous studies were missing two key components: a) having students self-reflect on assessment consistency and, b) completing exams in a real-world environment (a non-lab setting). This current study examines how participants self-rate their consistency when completing an assessment using real FAA Knowledge Exam questions for the Private Pilot Certificate. This research aims at assessing how consistent participants are in their test-taking, how consistent they think they are, and how well they self-assess their own consistency.

Test-Taking Performance

The process to help increase student performance has been thoroughly studied and there are various strategies employed to help enhance performance, including increasing knowledge of test material, refining study strategies, and/or improving test-taking skills. Craik and Tulving (1975) discovered that deep processing is linked with improved memory processing compared to shallow processing. Deep processing concentrates on the understanding and meaning of information. On the other hand, shallow processing gives minimal attention to the meaning of information. Tactics utilized to develop deep processing include repetitive testing on materials (Wheeler, Ewers, & Buonanno, 2003), active and organizational note taking (Peverly, Brobst,

Graham, & Shaw, 2003), and practice retrieval of information (Karpicke, 2009). Through these processes, there is an increase in retention and retrieval of information, which subsequently enhance test performance.

Increasing assessment performance may be linked to the results found in the improvement of student strategies (Fleming, 2002). By employing various study techniques, individuals may find improvements in assessment performance, even though different strategies may be better suited to different assessment structures. When a study strategy is not suitable for the assessment, there may be a decrease in test performance (Balch, 2007). One other advantage of utilizing a study strategy is the use of dedicated study time. When students set aside a distinct amount of time for studying, their performance tends to increase when contrasted to cramming for a test (Kornell, 2009; Smith & Rothkopf, 1984).

It has been found that some individuals are just better test-takers than others. The term test-wiseness is used to describe these types of individuals (Millman, Bishop, & Ebel, 1965; Rogers & Yang, 1996). These individuals tend to read test questions meticulously (Cohen, 2006), predict answer selection before reading available answers (McClain, 1983), and evaluate all options before making a decision. Good time management skills when completing an assessment, and making educated guesses while answering the questions that they are unsure of, are characteristics of a test-wise individual. Ostensibly, through these strategies, it is likely that they will lead to more consistency in their answers.

Consistency

Depending on the field of research, consistency is defined differently. Attribution research sees consistency as the similarity in response of individuals across different situations (Kelley & Michela, 1980; Orvis, Cunningham, & Kelley, 1975). Brunswik (1952), within the

Lens model, delineates consistency as the correlation between actual judgments and predicted judgments. For the purpose of this study, consistency refers to "when a person responds in an identical manner when presented with multiple identical situations." The correlation coefficient between two blocks of identical test questions is our operational definition. We refer to this value as the *consistency coefficient*.

Trafimow and Rice (2008; 2009) have studied the consistency coefficient where participants are given two similar blocks of questions. Based on participant responses, a within-persons correlation is completed, and this creates a quantifiable gauge of a participant's consistency. Prior research on human-automation performance, counting, educational test-taking, memory, time pressure, visual search, perception, aerial reconnaissance, and morality have used within-persons consistency coefficients (Hunt, Rice, Trafimow, & Sandry, in press; Rice, Geels, Hackett, et al., 2012; Rice, Geels, Trafimow, & Hackett, 2011; Rice & Trafimow, 2012a, 2012b; Rice, Trafimow, & Hunt, 2010; Rice, Trafimow, Keller, Hunt, & Geels, 2011; Rice, Trafimow & Kraemer, 2012; Trafimow, Hunt, Rice, & Geels, 2011; Trafimow, MacDonald, & Rice, 2012; Trafimow & Rice, 2008, 2009, 2011). One common trend throughout these studies was participants' rather low consistency, particularly when the task was difficult. These previous studies create a framework of using consistency coefficients as a tool for measuring assessment performance, and have established a need for further research on the role of consistency in performance.

Performance and student consistency appear to be linked. Using Potential Performance Theory (PPT), Rice, Geels, Trafimow, and Hackett (2011) examined student test-taking consistency. The study asked a sample of undergraduate students, divided into 13 subject areas, to complete a 50-question assessment. After a short period of time, students were asked to

complete the assessment again. The researchers discovered that a lack of consistency in answering the assessment led to a decrease in the participants' score by between 3%-20%. If students increased their response consistency, then many could subsequently improve their overall performance. For example, if some of the students who partook in this study had improved their consistency, it would have resulted in a perfect score on the assessment. This would translate into the possibility that these individuals would not need to spend more time gaining knowledge, but rather should spend time improving assessment consistency.

Cultural considerations may also have an influence on consistency of assessment score. Rice and Trafimow (2012b) studied two groups of participants from the United States and India, by giving each group a 50-question assessment, a short break, followed by the same assessment again. Indian participants scored 8% higher than participants from the United States and had much better consistency. From a systematic perspective, the Indian participants performed better. The American participants seemed to struggle with their consistency on the assessment, which may have resulted in a lower score.

In 2012, Rice, Trafimow, and Kraemer conducted a study comparing consistency and assessment improvement. Over the course of three sessions, the participants were given a true/false history assessment six times. Although the participants did show an increase in their consistency, the researchers also discovered that their systematic factors offset the increase in assessment outcome by consistency.

Current Study

While previous research has shown the importance of consistency in examinations, they did not look at how participants self-reflect their consistency when completing assessments. Another important gap is that the previous research was conducted in laboratory settings, and not

with real-world examinations. In the current study, we had students from the subject university complete two examinations (with two identical but randomized versions) at different points in the semester that involved a portion of the FAA Private Pilot Knowledge Exam for credit in their Aeronautics 1 (private pilot) course. Our hypotheses were as follows:

A) Based on previously completed studies (e.g. Rice, et al., 2011), it was predicted that participant's consistency coefficients (actual consistency between exams) would be low.

B) Participants were expected to rate themselves as being moderate to high in their self-assessment of consistency indicating that they thought they performed consistently.

C) Participants' correlation between their actual consistency score (consistency coefficient) and self-assessment score would be low, indications that participants are poor in accurately self-assessing their own consistency.

Methodology

Exam 1 - Method

Participants. Sixty-five (10 females) participants from the subject university took part in the study. The mean age was 18.95 (*SD* = 1.62).

Materials and procedure. Participants were given a sub-section of the FAA Private Pilot Knowledge Exam questions. A random selection of 30 questions was created using a third party online study systems. Questions were selected from the appropriate unit materials for which the exam covered. Following this, participants were given a short break and then presented with the same exam in a second block. Questions were randomized in each exam version; however, the actual questions were identical. The reason for presenting participants with the identical exam twice was to generate a correlation coefficient (we refer to this as the

consistency coefficient) between the two blocks of test questions; that is, we were interested in seeing how consistent participants were in answering identical questions at different times.

After answering the two blocks of test questions, participants were asked how consistently they thought they had answered the questions. Their answers were given on a 7-point Likert scale from extremely inconsistent (-3) to extremely consistent (+3), with a neutral option of zero. Lastly, participants were asked basic demographics questions, debriefed and dismissed.

Exam 1 – Results

First, we calculated a consistency coefficient for each participant between the two blocks of test questions. The average consistency coefficient for all participants was $r = .89$, indicating that, in general, participants were very consistent in their responses across the two blocks of test questions. Second, we calculated how consistent participants thought they were and found a relatively high self-evaluation ($M = 1.62$, $SD = 1.64$). Lastly, since each participant had a consistency coefficient as well as a self-rating of consistency, the researchers were able to calculate the correlation between their consistency coefficient and their self-ratings. This correlation was $r_s = .55$.

Exam 1 – Discussion

Although participants were fairly consistent across the two exams, and they were very confident about their consistency, the correlation between their consistency coefficient and their self-ratings were not impressive. In fact, these ratings only accounted for about 30% of the variance in the data. Thus, the findings suggest that, in general, participants were rather lacking in terms of judging their own levels of consistency. What is interesting, however, is the high consistency across the two exam versions. Given that all of the previous research in the field of

consistency indicates that a difficult task should result in lower consistency (Trafimow & Rice, 2009), the results were not as expected, which will be discussed further in later sections.

Exam 2 – Introduction

Exam 2 was conducted as a replication of Exam 1. Our goal was to provide a replication of the results in order to generate stronger external validity.

Exam 2 – Method

Participants. Fifty-five (7 females) participants from the subject university took part in the study. The mean age was 19.15 (SD = 1.65).

Materials and procedure. Exam 2 was identical to Exam 1 with the exception that the questions generated from the FAA databank were different.

Exam 2 – Results

The average consistency coefficient for all participants was r = .80, replicating the high consistency from the first exam. Their self-ratings were somewhat lower than in the first exam (M = 0.72, SD = 2.05); however, they still tended to believe they were more consistent than inconsistent. The correlation between their consistency coefficient and their self-ratings was very low at r_s = .11, which accounts for only 1% of the variance in the data.

Exam 2 – Discussion

The results in Exam 2 were generally similar to those obtained in Study 1. Participants' consistency coefficients were again surprisingly high. More details about these results are found in the following section. Participants' self-ratings of consistency appeared to drop in Exam 2, which may be due to a learning effect, but still higher than neutral. The very low correlation between their consistency coefficients and their self-ratings of consistency indicated that, in

general, participants were very poor at determining their own test-taking consistency. Table 1 provides a summary of results for the two exams.

General Discussion

The purpose of the current study was to examine student exam-taking consistency when completing a real-world examination. The researchers also sought to determine how well participants would self-assess their consistency when completing identical examinations, and to evaluate the relationship between the participants' actual consistency coefficient and their self-assessment score. Consistency has been shown to be valuable in optimizing student performance when completing assessments (Rice, Geels, Trafimow, & Hackett, 2011). A unique aspect to the current study was the use of a real-world exam, whereas previous studies have been conducted in lab settings (Hunt, Rice, Trafimow, & Sandry, in press; Rice, Geels, Hackett, et al., 2012; Rice, Trafimow, & Hunt, 2010; Rice, Trafimow, & Kraemer, 2012; Sandry et al., in press). The current study utilized questions from the Federal Aviation Administration's Private Pilot Knowledge Exam and was completed in a real-world setting where participants were studying to complete the official examination for pilot certification.

The first hypothesis was that participants' consistency coefficient would be low based on the findings of previous studies. However, data from the study refuted this hypothesis that participants' consistency coefficient (actual consistency correlation between exams) would be low. On the contrary, participants' actual correlation between exams was quite high for Exam 1 and 2, $r = .89$, and $r = .80$, respectively. This finding is significant as it is the opposite of the results in earlier work on this topic. There may be a few possible explanations for this difference. This study, opposed to previous ones, was completed in a real-world setting. While previous studies were completed in lab settings, it is possible that participants did not feel the

need to perform to the best of their ability because there was no motivating factor for successful completion. However, in the current study, participants were working toward a real-world goal of successfully passing an FAA examination, along with obtaining a grade in the course. Additionally, participants also studied and prepared for the exam, while in previous studies, participants were administered questions without any preparation time.

The second hypothesis anticipated participants' self-assessment of their consistency would be moderate to high. The data mostly supported this finding with participants' self-assessing their consistency on Exam 1 (M = 1.62, SD = 1.64) and Exam 2 (M = 0.72, SD = 2.05) as above neutral. Perhaps initially, participants were over-confident in their consistency level as participants did have a slight decrease in their self-assessment score of consistency on Exam 2. It is possible that participants were more aware of consistency on Exam 2, and perhaps were more cautious and thoughtful in their response on the second exam. Participants may have also experienced a learning effect, and learned something about themselves from the previous exam, which may have assisted in providing a more calculated rating of their self-assessment on consistency in completing Exam 2.

In the third hypothesis, it was predicted that the correlation between the consistency coefficient and participants' self-rating of consistency would be low. This relationship was r_s = .55 and r_s = .11 on Exams 1 and 2, respectively. There is a noticeable drop in scores between Exams 1 and 2, and this seems to support the hypothesis that participants would have very low predictive power of self-assessment. However, this finding is not much of a surprise. Previous studies have shown that people frequently do poorly on self-assessment tasks (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger & Dunning, 1999). Additionally, Kruger and Dunning (1999) cite a double curse when participants try to complete self-assessment: 1) participants may

not have the knowledge to perform well on the exam, and 2) they then also lack the knowledge to realize how poorly they are performing. Unfortunately, this often results in a skewed or inaccurate self-assessment of how participants actually did on the task.

Practical Implications and Limitations

The findings from this study have certain practical implications. First, this study was completed in a real-world setting using actual questions from the FAA's Private Pilot Knowledge Exam bank of questions. Participants were working toward obtaining their pilot certifications and achieving a successful grade in the collegiate level class. These factors could provide motivation for increased performance and better consistency. Previous research has demonstrated that it is not only important to have the proper knowledge to perform well on an exam, but improvements in consistency also work to optimize student performance (Rice, Geels, Trafimow, & Hackett, 2011).

It has been mathematically proven that inconsistency reduces overall performance (Trafimow & Rice, 2008), as long as the person is performing at a rate that is better than random. There are cases where a person might perform at worse than random, but these cases are highly unlikely given the nature of the FAA exam. Thus, in practical terms, anytime a student is less than perfectly consistent, they are performing at a lower accuracy than they should be. Potential Performance Theory (PPT) by Trafimow & Rice (2009) reveals how to assess what a person's potential performance would be given perfect accuracy; however, to date, that theory only deals with dichotomous data (i.e. true-false, yes-no, etc.). Given that the data in the current study is not dichotomous, the authors could not assess what a person's potential performance would be given perfect consistency. With that limitation, our findings show that many of the participants were less than perfectly consistent, and therefore, did not perform at their optimal level. Future

research could both test the same phenomenon using dichotomous data, and determine how PPT could apply to this type of data.

This study is especially applicable in the field of aviation education research. To obtain many of the FAA's pilot certifications, the successful completion of a knowledge or written examination is required. Participants have study materials available to them from a number of commercial, third party sources. While it is important that they study the material to increase their knowledge levels, it is also important that participants answer similar questions consistently. For example, the Private Pilot Knowledge Exam test bank consists of around 700 questions total, but participants only receive a random sample of 60 on their actual exam. Therefore, for each subject area, there may be multiple, similarly phrased questions, of which the participant may only receive one on the actual examinations.

While studying, it is likely that participants will study all variations of these questions. If aware of consistency, participants can apply a consistent response method to all questions from that section, hopefully reducing the amount of study time required and making the process more efficient. Answering consistently may also improve student confidence due to increased performance and scores on practice examinations. Future studies should continue to examine this line of research in real-world settings to see if there is a relationship between consistency and confidence in test taking.

As with all studies, the current study is not without certain limitations. Only students from the subject university's private pilot equivalent ground school were examined. Further research should expand this sample to include other private pilot ground schools in both university and non-university environments; this also limits the generalizability of the findings. Additionally, this is one of the first studies to examine this research track in a real-world setting.

Replicating this type of study in other real-world settings could help to verify the findings of this study.

Conclusions

This study reviewed the consistency of participants when completing an exam in a real-world setting. Participants had high consistency coefficients when completing two versions of the same exam. This is different from previous research findings in the field of student consistency on examinations, and possibly explained by the use of a real-world FAA exam. Participants also had moderate self-assessments of their own consistency, which decreased slightly on the second exam, perhaps indicating a learning effect on consistency had occurred. Finally, the correlation between actual consistency and self-assessment of consistency was fairly low, especially on the second exam, which suggests that participants have very low predictive ability to accurately self-assess their consistency. This finding is similar to other studies related to participant self-assessment. Consistency has been found to play a role in performance, and this line of research can drastically impact the understanding of student performance in examination settings.

Acknowledgments

The authors wish to thank the 1st annual Aviation, Aeronautics, and Aerospace International Research (A3IR) Conference held January 17th-18th, 2014 in Phoenix, AZ and hosted by Embry-Riddle Aeronautical University, where much of this data was presented. Additionally, special thanks are given to the faculty of the College of Aeronautics at the Florida Institute of Technology for their help in the data collection process.

References

Balch, W. R. (2007). Effects of test expectation on multiple-choice performance and subjective ratings. *Teaching of Psychology, 34*(4), 219-225. http://dx.doi.org/10.1080/00986280701700094

Brunswik, E. (1952). *The conceptual framework of psychology: International encyclopedia of unified science.* Chicago: University of Chicago Press.

Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Learning Assessment Quarterly, 3*(4), 307-331. http://dx.doi.org/10.1080/15434300701333129

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*(3), 268-294. http://dx.doi.org/10.1037/0096-3445.104.3.268

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83-87. http://dx.doi.org/10.1111/1467-8721.01235

Fleming, V. M. (2002). Improving students' exam performance by introducing study strategies and goal setting. *Teaching of Psychology, 29*(2), 115-119. http://dx.doi.org/10.1207/S15328023TOP2902_07

Hunt, G., Rice, S., Trafimow, D., & Sandry, J. (in press). Using Potential Performance Theory to analyze systematic and random factors in enumeration tasks. *American Journal of Psychology*.

Karpicke, J. D. (2009). Metacognitive control strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology, 138*(4), 469-486. http://dx.doi.org/10.1037/a0017341

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology, 31,* 457-501. http://dx.doi.org/10.1146/annurev.ps.31.020180.002325

Kornell, N. (2009). Optimizing learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*(9), 1297-1317. http://dx.doi.org/10.1002/acp.1537

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. http://dx.doi.org/10.1037/0022-3514.77.6.1121

Kubiszyn, T., & Borich, G. (2010). *Educational testing & measurement: Classroom application and practice.* Hoboken, John Wiley & Sons, Inc.

McClain, L. (1983). Behavior during examinations: A comparison of "A," "C," and "F" students. *Teaching of Psychology, 10*(2), 69-71. http://dx.doi.org/10.1207/s15328023top1002_2

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement, 25*(3), 707-726. http://dx.doi.org/10.1177/001316446502500304

Orvis, B. R., Cunningham, J. D., & Kelley, H. H. (1975). A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology, 32*(4), 605-616. http://dx.doi.org/10.1037/0022-3514.32.4.605

Peverly, S. T., Brobst, K. E., Graham, M., & Shaw, R. (2003). College adults are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking. *Journal of Educational Psychology, 95*(2), 235-346. http://dx.doi.org/10.1037/0022-0663.95.2.335

Rice, S., Geels, K., Hackett, H., Trafimow, D., McCarley, J. S., Schwark, J., & Hunt, G. (2012). The harder the task, the more inconsistent the performance: A PPT analysis on task difficulty. *Journal of General Psychology, 139*(1), 1-18. http://dx.doi.org/10.1080/00221309.2011.619223

Rice, S., Geels, K., Trafimow, D., & Hackett, H. (2011). Our students suffer from both lack of knowledge and consistency: A PPT analysis of test-taking. *US-China Education Review, 1*(6), 845-855.

Rice, S., & Trafimow, D. (2012a). Time pressure heuristics can improve performance due to increased consistency: A PPT methodology. *Journal of General Psychology, 139*(4), 273-288. http://dx.doi.org/10.1080/00221309.2012.705187

Rice, S., & Trafimow, D. (2012b). Using Potential Performance Theory to assess differences in math abilities between citizens from India and the United States. *Higher Education Studies, 2*(3), 24-29. http://dx.doi.org/10.5539/hes.v2n3p24

Rice, S., Trafimow, D., & Hunt, G. (2010). Using PPT to analyze sub-optimal human-automation performance. *Journal of General Psychology, 137*(03), pp. 310 - 329. http://dx.doi.org/10.1080/00221301003645236

Rice, S., Trafimow, D., Keller, D., Hunt, G., & Geels, K. (2011). Using PPT to correct for inconsistency in a speeded task. *Journal of General Psychology, 138*(1), 12-34. http://dx.doi.org/10.1080/00221309.2010.531791

Rice, S., Trafimow, D., & Kraemer, K. (2012). Using Potential Performance Theory to assess how to increase student consistency in taking exams. *Higher Education Studies, 2*(4) 68-74. http://dx.doi.org/10.5539/hes.v2n4p68

Rogers, W. T., & Yang, P. (1996). Test-wiseness: Its nature and application. *European Journal of Psychological Assessment 12*(3), 247-259. http://dx.doi.org/10.1027/1015-5759.12.3.247

Sandry, J., Rice, S., Trafimow, D., Hunt, G., Busche, L., & Rubio, E. (in press). Suboptimal recognition for increasing set sizes is largely due to inconsistency: A potential performance theory analysis of individual differences. *Cognitive Technology*.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*(3), 341-358. http://dx.doi.org/10.1207/s1532690xci0103_4

Trafimow, D., Hunt, G., Rice, S., & Geels, K. (2011). Using potential performance theory to test five hypotheses about meta-attribution. *Journal of General Psychology, 138*(2), 1-13. http://dx.doi.org/10.1080/00221309.2010.540591

Trafimow, D., MacDonald, J., & Rice, S. (2012). Using PPT to account for randomness in perception. *Attention, Perception and Psychophysics, 74*, 1355-1365. http://dx.doi.org/10.3758/s13414-012-0319-7

Trafimow, D., & Rice, S. (2008). Potential Performance Theory: A general theory of task performance applied to morality. *Psychological Review, 115*(2), 447-462. http://dx.doi.org/10.1037/0033-295X.115.2.447

Trafimow, D., & Rice, S. (2009). Potential Performance Theory (PPT): Describing a methodology for analyzing task performance. *Behavior Research Methods, 41*(2), 359-371. http://dx.doi.org/10.3758/BRM.41.2.359

Trafimow, D., & Rice, S. (2011). Using a sharp instrument to parse apart strategy and consistency: An evaluation of PPT and its assumptions. *Journal of General Psychology, 138*(3), 169-184. http://dx.doi.org/10.1080/00221309.2011.574173

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571-580. http://dx.doi.org/10.1080/09658210244000414

Table 1

Summary of Results

Instrument	Consistency Coefficient	Self-Assessment Ratings	Consistency Coefficient and Self-Assessment Correlation
Exam 1	$r = .89$	$M = 1.62$ $SD = 1.64$	$r_s = .55$
Exam 2	$r = .80$	$M = 0.72$ $SD = 2.05$	$r_s = .11$

https://commons.erau.edu/jaaer/vol24/iss2/2
DOI: https://doi.org/10.15394/jaaer.2015.1608