

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 27 Number 8, April 2022

ISSN 1531-7714

Reinstatement Candidate Credentialing Exam Performance: Evaluating the Persistence of Misinformed Responses on Multiple Choice Items¹

Ben Babcock, *Elsevier*

Zachary D. Siegel, *The American Registry of Radiologic Technologists*

Research about repeated testing has revealed that retaking the same exam form generally does not advantage or disadvantage failing candidates in selected response-style credentialing exams. Feinberg, Raymond, and Haist (2015) found a contributing factor to this phenomenon: people answering items incorrectly on both attempts give the same incorrect response about 2/3 of the time. They concluded that examinees are misinformed, rather than uninformed, about these items. The current research investigated whether reinstatement candidates followed similar patterns. Reinstatement candidates are people that obtain a credential, later discontinue the credential, then retake the exam to regain the credential. Data came from a major certification exam program in medical imaging. Candidates' reinstatement attempts had questions in common with their earlier passing attempts. Results showed that, similar to Feinberg et al., candidates answering questions incorrectly on both passing and reinstatement attempts gave the same incorrect response 65.7% of the time. It appears that professional misconceptions are persistent for numerous years. Other patterns of correct and incorrect responses were consistent when considering the results of both Feinberg et al. and recent research on reinstatement candidates. Results concerning changes in the time spent on each question, however, were different from Feinberg et al. The current study found no substantial patterns in response time change between subsequent attempts for items seen previously. This could have to do with the fact that the items in common between the two exam attempts were only a portion of the larger exam form.

Introduction

Extensive research has investigated the effects of giving credentialing exam candidates the same versus a different parallel form for a retake attempt of a selected response exam. Results have largely found that, for candidates failing a first attempt, receiving the same scored form does not greatly advantage or

disadvantage retake candidates (Geving, Webb, & Davis, 2005; O'Neill, Sun, Peabody, & Royal, 2015; Raymond, Neustel, & Anderson, 2007; Raymond, Neustel, & Anderson, 2009; Wood, 2009). These results have been useful to boards crafting exam retake policies. Such findings are also useful in making form publication guidelines for small-sample programs, as gathering additional data by republishing a scored

¹ The majority of the work for this research was conducted while both authors were employed by The American Registry of Radiologic Technologists. The views and discussions contained in this paper are solely the authors' and are not necessarily the official positions of ARRT or Elsevier.

exam form can be statistically advantageous for item response theory-style linking and equating (Babcock & Hodge, 2020).

More recent research by Feinberg, Raymond, and Haist (2015; herein referred to as Feinberg et al.) investigated item-level data to answer why repeat candidates are not advantaged by taking the same form versus a parallel form. Results showed that more items changed from incorrect to correct on the second attempt compared to the number of items that went from correct to incorrect. Candidates had a decrease in response time for the incorrect to correct items, whereas those same candidates had an increase in time spent on the correct to incorrect items. This indicates that candidates may indeed remember and study certain items, but the effects were not advantageous compared the score gains observed with using parallel forms.

One of the most interesting findings of Feinberg et al., however, was their analysis of items that candidates marked incorrectly on both attempts. The researchers found that candidates responding incorrectly to the same item twice picked the same incorrect distractor in 68% of cases. The chances of picking the same incorrect distractor with a random guessing mechanism from a four-option multiple-choice item (removing the correct option as a possibility) are only 33%. The fact that the observed percentage is substantially greater than chance led the authors to conclude that, for many items, exam candidates harbor misconceptions that result in repeating incorrect answers. This phenomenon may be a contributor as to why repeating the same form does not seem to advantage candidates in a credentialing context.

Feinberg et al. revealed excellent insights into retesting behavior and is one of the best studies to date on the topic. Just as with any study, there were some questions that were not answered by Feinberg et al. Two questions concerning credentialing exam retake behavior had to do with score distributions and lag time between attempts. First, because Feinberg et al. focused on initial credentialing retake candidates, the study only included people with low initial attempt scores. This is expected, as most boards' retake policies prevent passing candidates from retaking an exam. It is unclear whether the patterns found in their research

will be replicable in a population that has higher scores on the initial attempt studied. Second, the exam attempts for Feinberg et al. were one year apart. It is unclear whether the pattern of harboring misconceptions will persist for longer amounts of time.

Past research in a medical certification context has also shown that exam performance generally declines over time (Leigh, Young, & Haley, 1993). Past research has further shown that there can be a general degradation in factual knowledge in many individuals as they age (Brannon, Koubeck, & Voss, 2008). It is unclear whether reinstatement candidates will follow these same patterns, wherein longer lag times between exams will correlate with lower reinstatement exam performance.

Reinstatement Candidates

Reinstatement candidates can provide researchers with the opportunity to examine both whether misconceptions persist for longer amounts of time and to see whether other trends in repeat response behavior are similar with a population that scored higher on an earlier attempt. Reinstatement candidates successfully achieve a credential, discontinue that credential, then later retake an examination to regain status as credentialed. People may discontinue credentials for a variety of reasons, including changing careers, failing to comply with renewal requirements, or having a credential revoked based on an ethics violation (Freed, Abraham, & Brzoznowski, 2007). These candidates could, in theory, share some exam items between their passing attempts and reinstatement attempts.

Relatively little has been published in the peer-reviewed literature about the performance of reinstatement candidates. This is not surprising, as many credentialing programs do not have enough reinstatement candidates for a meaningful study. In one of the few studies on the topic, Babcock (2021) found that reinstatement candidates performed substantially worse than both initial certification candidates and the reinstatement candidates' own passing attempt. While these performance differences varied in degree by content, reinstatement candidates performed worse in every section of content for the exam studied. Reinstatement candidates did, however, perform better than failing initial exam candidates taking subsequent attempts.

Current Study

We will conduct similar analyses to those by Feinberg et al. for reinstatement candidates. There were two research questions that we sought to answer. First, what are the patterns in the same item responses for reinstatement candidates compared to the candidates passing attempt? Second, for items where reinstatement candidates respond incorrectly on both attempts, what percentage of responses were the same incorrect answer on both attempts?

Method

Data

Data in this study came from 683 reinstatement candidates from a major certification program in medical imaging. This certification program is accredited by the NCCA (Institute for Credentialing Excellence, 2020) for adhering to high-quality credentialing standards, including those for testing and measurement. Candidates took the exam as reinstatement candidates between January 1, 2016 and May 5, 2020. All candidates had an earlier passing attempt on or after the year 2000, which is when the certification program began administering exams with computer delivery. The median lag time between the passing attempt and the reinstatement attempt was 9.8 years. Table 1 contains additional descriptive statistics about lag time.

The certification exam had 200 scored items. We eliminated items that were of response types other than multiple choice, such as sorted list and select multiple, from this analysis to maintain comparability with Feinberg et al. Because of the significant time lag between passing and reinstatement attempts, only some of the items on the exam were the same. The median number of items shared between the two attempts was 15. While this is not a great number of responses for an individual, combining across all candidates yielded 10,749 responses. Table 1 contains additional descriptive statistics about the items in common between the passing and reinstatement attempts.

In order to see how the exam program in question compared to the program studied by Feinberg et al., we also conducted all analyses using first-time and repeat

candidates (i.e., non-reinstatement candidates). A total of 4,737 repeat candidates took the exam between January 1, 2016 and May 5, 2020. These candidates shared a total of 132,041 responses between their first attempts and repeat attempts. While not the central focus of this study, analyses of this group will help create a context for the results of the reinstatement candidates.

Analyses

The first set of analyses will be general analyses of reinstatement and first-time candidate performance. First, we compared the reinstatement candidates' initial passing attempt performance to the performance of other test takers in that same time period. The purpose of this analysis is to see whether reinstatement candidates' base performance is typical or atypical of all exam candidates. Second, we conducted analyses of exam performance and the lag time between the initial passing attempt and reinstatement attempt. These analyses will help to see whether the amount of raw lag time is a significant factor in exam performance (as in past studies).

The second set of analyses, which is the main focus of this study, will concentrate on individual answers. First, we classified responses into four response pattern categories based on whether candidates responded as correct or incorrect on the earlier attempt and correct or incorrect on the later attempt. Second, we calculated the mean and standard deviation of the change in response time for the items in each of these response categories. Third, for only those responses where there was an incorrect answer on both attempts, we calculated the percentage of responses where candidates marked the same incorrect answer on both attempts. Finally, for the reinstatement candidates only, we calculated a Pearson correlation to see if the amount of time lag correlated with the proportion of incorrect answers selecting the same response on both attempts.

Results

Comparing Reinstatement Candidates' Initial Passing Attempt to Other Exam Candidates

Table 2 contains some basic descriptive statistics on scaled score for initial exam attempts during the time period when the reinstatement candidates passed

Table 1. Basic Descriptive Statistics for Reinstatement Candidates

N Candidates	683	
N Total Responses, Both Attempts	10,749	
Statistic	Lag between Passing and Reinstatement Attempts (Years)	N Items in Common between Passing and Reinstatement Attempt
Min	2.8	2
First Quartile	6.9	12
Median	9.8	15
Third Quartile	13.1	19
Max	19.5	34

Table 2. Descriptive Statistics for Overall Scaled Score Exam Performance for Initial Attempts

Population	Mean	SD	First Quartile	Median	Third Quartile
All Exam Attempts	82.40	8.25	77	84	89
First-Time Candidates' Attempts	84.13	7.30	80	85	89
Reinstatement Candidates' Initial Passing Attempt	84.25	5.47	80	84	89

the initial exam. As one can see, the mean and quartiles are very similar for the reinstatement candidates' initial passing attempt and the performance of all first-time candidates. The reinstatement candidates' initial passing attempt scores were generally a bit higher than all exam attempts. The standard deviation of the reinstatement candidates' exams was somewhat lower than the first-time candidates because the reinstatement candidate scores only included passing scores. We also conducted an analysis of reinstatement candidates' first attempts (not included in the table). There was not much of an effect on the statistics displayed here; around 90% of the reinstatement candidates had indeed passed on their first initial exam attempts.

Analysis of Time Lag and Overall Reinstatement Performance

We calculated correlations for the time lag between passing and reinstatement attempts with both the reinstatement score and the change in score between the initial and remediation attempt. Both correlations had an absolute value of less than .05 and were not statistically significant using any typical p-value cutoff. In order to demonstrate that there were no nonlinear effects, Figure 1 contains a scatterplot of scaled score change and lag time.

In order to see if there were any more complex effects, we also conducted a linear regression using initial passing score, lag time, and the interaction between initial passing score and lag time to predict the score on the reinstatement attempt. While initial passing score was, unsurprisingly, a significant predictor of reinstatement score, neither the main effect nor the interaction involving lag time were significant predictors of reinstatement score.

Individual Answers: Repeat Candidate Performance (Non-Reinstatement)

Table 3 contains the analysis of repeat candidates. The percentage of responses in each of the four response pattern categories was similar to Feinberg et al. The largest proportion of responses was both correct, followed by both incorrect, followed by incorrect to correct, and last by correct to incorrect. The net performance increase found by Feinberg et al. (percent incorrect to correct minus percent correct to incorrect) was 4.7%. This study's net change was within 1% of Feinberg et al. at 5.6% net improvement. Among the items incorrect on both attempts, Feinberg et al. found that around 68% of the incorrect responses were the same distractor on both attempts. The repeat candidates in this study responding incorrect on both

Figure 1. Plot of Scaled Score Change on Reinstatement Lag Time

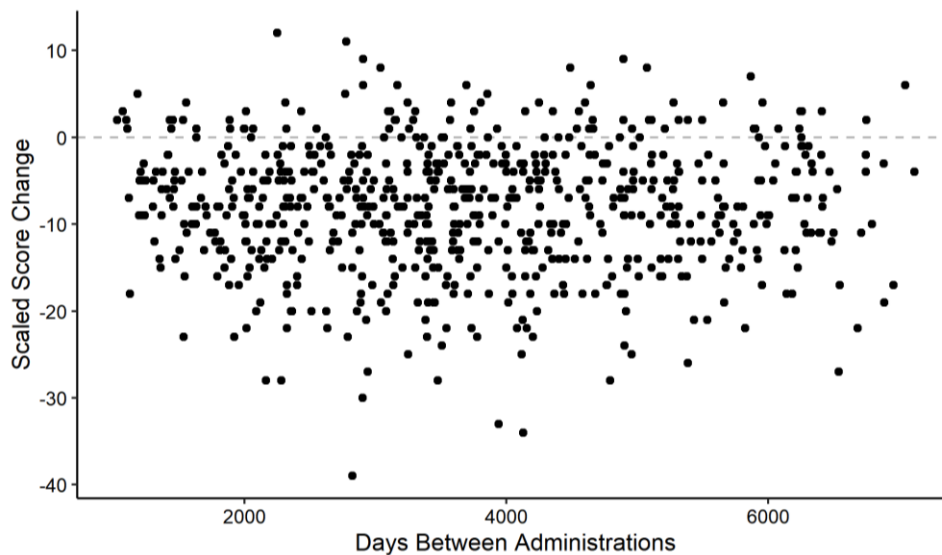


Table 3. Regular Repeat Candidate Same Item Response Patterns and Change in Response Time

Same Item Response Pattern (Passing First, Reinstatement Second)	Percent	Change in response time (seconds)	
		M	SD
Correct, Correct	50.2	0.59	40.71
Correct, Incorrect	11.0	0.58	40.36
Incorrect, Correct	16.6	0.62	38.77
Incorrect, Incorrect	22.2	0.88	38.03

Table 4. Reinstatement Candidate Same Item Response Patterns and Change in Response Time (Same Analyses as Conducted by Feinberg et al.)

Same Item Response Pattern (Passing First, Reinstatement Second)	Percent	Change in response time (seconds)	
		M	SD
Correct, Correct	63.9	-0.69	38.25
Correct, Incorrect	17.2	-1.32	42.25
Incorrect, Correct	9.3	2.43	37.77
Incorrect, Incorrect	9.5	-0.39	37.07

attempts gave the same incorrect answer 67.8% of the time.

The response time results were less similar. Feinberg et al. found meaningful changes in response time between the attempts; the repeat candidates in this study did not appear to have major differences in response time. The mean change in response times in this study were all less than one second, and the standard deviations of the change in time were all close to 40 seconds.

Individual Answers: Reinstatement Candidate Performance

Table 4 contains the results of the response pattern analysis. Consistent with Feinberg et al., the largest percentage of responses was the category in which candidates marked a correct response on both attempts. The current study found a larger percentage of items in this category than Feinberg et al. That is an expected result given that the first time point for responses was a passing attempt in this study instead

of a failing attempt in Feinberg et al. Also relatively consistent with Feinberg et al. was that the percentage of items going from incorrect to correct was relatively close to the percentage of items that were incorrect on both attempts.

There were, however, some differences between the Feinberg et al. results and the current reinstatement results. The first difference is with the relative percentages of items that changed from correct to incorrect or incorrect to correct on subsequent attempts. Feinberg et al. found that there were more items that changed from incorrect to correct than that changed from correct to incorrect. The difference amounted to a 4.7% increase in overall percent correct. In contrast, this study found that, among reinstatement candidates, there were more items that went from correct on the passing attempt to incorrect on the reinstatement attempt rather than incorrect to correct. The difference amounted to a 7.9% decrease in exam performance.

The second difference is with response times. Feinberg et al. found that items answered correctly on the second attempt tended to have slightly faster response times, and items answered incorrectly on the second attempt had slightly slower response times. In contrast, this study found very little change in mean response time based on whether an item was correct or incorrect. In fact, the greatest difference in mean response time found for reinstatement candidates (incorrect to correct items) was close in absolute value to the smallest mean response time difference from Feinberg et al. (correct to correct items). The standard deviations in response time change were also substantially lower in the current study and relatively similar across all response pattern categories.

As a final set of analyses, this study examined the responses that were incorrect on both attempts to see what percentage of people responded incorrectly with the same response option. This study found that 65.7% of responses that were incorrect on both passing and reinstatement attempts had the same incorrect answer. This percentage is quite close to the 68% found by Feinberg et al. The correlation between the percentage of same incorrect responses and time lag between passing and reinstatement attempts was not significantly different from zero ($\alpha = .01$) when using all candidates and when using only candidates with 5 or more incorrect responses on both attempts. This

result is consistent with Babcock (2021), which found that time lag between passing and reinstatement attempts was not strongly related to reinstatement performance.

Discussion

This research sought to extend the work of Feinberg et al. by examining item-level data for reinstatement candidates on credentialing exams taking the same items after a time lag between passing and reinstatement attempts. An analysis of non-reinstatement repeat candidates showed that the correct and incorrect response patterns were relatively consistent to those reported in Feinberg et al., demonstrating that the certification exam programs were comparable in performance. It also demonstrates that the correct/incorrect performance results in Feinberg et al. generalize to other close-in-time initial certification retake contexts.

Initial analyses of the data revealed two points of context for this study. First, the initial passing exam performance of reinstatement candidates was comparable to first-time candidate performance. While the circumstances around their certification status make them quite different, it appears that reinstatement candidates are quite similar to candidates at large when it comes to initial certification exam performance. Second, the time lag between the initial and reinstatement attempt did not appear to be an influential factor on reinstatement score. While this finding is contrary to past research concerning time lag and exam performance (e.g., Leigh, Young, & Haley, 1993; Brannon, Koubeck, & Voss, 2008), it is consistent with past research on reinstatement candidates (Babcock, 2021). There are numerous potential possibilities for this finding, including the fact that lag time between exams is not the same as the amount of time out of practice. One person with an exam lag time of six years could have spent four of those six years out of practice, while another person with an exam lag time of 20 years may have only been out of practice for a few months due to issues such as continuing education compliance.

Analyses also showed that several interesting results from Feinberg et al. replicated in the reinstatement population. First, the largest percentage of items were those that people responded to correctly

on both attempts. This result is indicative that, whether exam attempts are relatively close in time for initial certification or separated by a substantial amount of time as reinstatement attempts, people that are qualified to take credentialing exams have and maintain a base level of construct-relevant knowledge. Second, the percentage of items that changed from incorrect to correct was relatively close to the percentage of items that remained incorrect on both attempts. A chance response mechanism for four-option multiple choice items would dictate that this ratio be closer to three-to-one in favor of both incorrect.

Finally, among those items that candidates responded to incorrectly on both attempts, around 2/3 of the incorrect responses in both studies used the same incorrect option. This is one of the most interesting findings from Feinberg et al., and it is remarkable that it generalizes to the reinstatement population. Feinberg et al. suggested that this large proportion of responding incorrectly twice using the same distractor indicated that candidates have misconceptions on certain topics that drive them to respond incorrectly in systematic ways. The phenomenon appearing in a reinstatement population pushes that conclusion further to suggest that some of these misconceptions can persist for a decade or even longer, even among higher-scoring candidates. Credentialing bodies could consider policy implications for maintaining a credential, such as targeted continuing education or continued assessment, given that these perceptions can persist for a long time. Such initiatives could help maintain and enhance the knowledge base of those certified and registered.

There were, however, some differences between results found by Feinberg et al. and those produced here in a reinstatement candidate population. First, reinstatement candidates had a net decrease in performance of 7.7%. Feinberg et al. found an increase in repeat candidate performance of 4.7%. This large difference in performance could be due to numerous factors that affect reinstatement candidates but not initial certification repeat candidates. Factors could include a larger time lag between administrations, a comparative lack of available study resources, and the amount of time potentially spent out of the job role. A limitation of the current study is that there were no data available about either the reason for people

discontinuing or the amount of time that people spent not working in the given discipline. Past research has shown that the reason for discontinuing may influence reinstatement exam performance (Cain, Davignon, Henzel, Ciccone, & Young, 2014). Furthermore, the reinstatement population always had a passing attempt as the past comparison exam attempt. From a statistical standpoint, it may be more difficult for higher-scoring individuals to maintain those scores on a subsequent attempt compared to lower-scoring failing candidates.

Another key difference was in the findings with response time. While Feinberg et al. found meaningful differences in response time based on the pattern of correct versus incorrect on the two attempts, this study found little in the way of response time patterns in either reinstatement candidates or repeat candidates. While there could be numerous reasons for this, we propose that one key difference in the design of Feinberg et al. and the current study may be driving the different response time results. The results analyzed in Feinberg et al. consisted of repeat candidates taking an intact set of scored multiple choice items. The current study harvested the selected items that two exams had in common, which was less than 20% of the test. Having a smaller number of previously seen items embedded among a longer test may make it more difficult for candidates to recognize that they have seen the items before compared to seeing all of the same items again.

As with any study, this study has some limitations. We will discuss four. First, this study examined smaller numbers of items in common from among a longer exam. Results could have differed if the reinstatement candidates had taken the same 200 item exam form twice. Such a study may be impossible to conduct, however, as there tends to be a substantial time lag between passing and reinstatement attempts. Changes in the credentialing program's content specifications over such long periods of time would necessitate that reinstatement candidates take a different exam form. Second, as previously stated, there were no data available concerning the reasons for candidates discontinuing their credentials. Data were also unavailable concerning how long reinstatement candidates were out of practice in medical imaging. These sorts of data could reveal valuable insights into how reinstatement candidates respond to previously viewed items. Future research should more thoroughly

explore such issues. Third, due to the timing of exams transitioning from paper-and-pencil to computer-based testing, the year 2000 was as far back in time as the passing attempts could go. That limited the largest possible time lag to around 20 years, with 75% of candidates having a time lag of 13.1 years or less. It is possible that the effects found in this paper could be different for longer time lags between passing and reinstatement attempts, such as 30 or more years. Finally, due to the fact that reinstatement candidates were both from a different population and scored higher than previous research populations on their passing attempts, it is somewhat unclear which effects may be due to candidates being higher scorers versus being reinstatement candidates per se. A retake study looking at candidates taking the same items for maintaining a credential (often called recertification) could provide insights. A retake study using higher scoring initial certification candidates retaking the same test items could also reveal this. This second type of study is, however, a study that is probably infeasible for reasons of exam security.

This study examined the correct and incorrect answer behavior among items that reinstatement candidates had viewed on a previous passing attempt. Results showed a pattern of correct and incorrect responses that is consistent with past research when combining the results of Feinberg et al. and Babcock (2021). Results also showed that, like Feinberg et al., around two thirds of the items where reinstatement candidates respond incorrectly twice are incorrect with the same distractor. This high percentage highlights that job-related misperceptions can continue for one or more decades. The persistence of these misperceptions underscores the importance for continuing education, assessment at multiple time points, or other initiatives that credentialing programs can implement to enhance the practice of their constituents.

References

- Babcock, B. (2021). Exam performance of reinstatement candidates pursuing ARRT certification and registration in radiography. *Radiologic Technology*, 92, 240-248.
<http://www.radiologictechnology.org/content/92/3/240.abstract>
- Babcock, B., & Hodge, K. J. (2020). Rasch versus classical equating in the context of small sample sizes. *Educational and Psychological Measurement*, 80, 499-521.
<https://doi.org/10.1177%2F0013164419878483>
- Brannon, N. G., Koubeck, R. J., & Voss, D. (2008). Mechanisms of knowledge degradation in a resource management task. *Theoretical Issues in Ergonomics Science*, 9, 25-44.
<https://doi.org/10.1080/14639220600731248>
- Cain, F. E., Davignon, P., Henzel, T. R., Ciccone, A., & Young, A. (2014). The special purpose examination: An evaluation of physicians taking SPEX and their pass rates. *Journal of Medical Regulation*, 100, 9-14. <https://doi.org/10.30770/2572-1852-100.1.9>
- Feinberg, R. A., Raymond, M. R., & Haist, S. A. (2015). Repeat testing effects on credentialing exams: Are repeaters misinformed or uninformed? *Educational Measurement: Issues and Practice*, 34, 34-39.
<https://doi.org/10.1111/emip.12059>
- Freed, G. L., Abraham, L. M., & Brzoznowski, K. F. (2007). Inactive physicians: The state of our understanding. *Journal of Pediatrics*, 151, 431-434.
<https://doi.org/10.1016/j.jpeds.2007.04.066>
- Geving, A. M., Webb, S., & Davis, B. (2005). Opportunities for repeat testing: Practice doesn't always make perfect. *Applied H.R.M. Research*, 10, 47-55. Retrieved online from:
http://applyhrm.asp.radford.edu/Volume%2010/MS%2010_2_%20Geving.pdf
- Institute for Credentialing Excellence. (2020). *NCCA accreditation*. Retrieved from
<https://www.credentialingexcellence.org/ncca>.
- Leigh, T. M., Young, P. R., & Haley, J. V. (1993). Performances of family practice diplomates on successive mandatory recertification examinations. *Academic Medicine: Journal of the Association of American Medical Colleges*, 68, 912-919.
<https://doi.org/10.1097/00001888-199312000-00012>
- O'Neill, T. R., Sun, L., Peabody, M. R., & Royal, K. (2015). The impact of repeated exposure to items. *Teaching and Learning in Medicine*, 27, 404-409.
<https://doi.org/10.1080/10401334.2015.1077131>
- Raymond, M. R., Neustel, S., & Anderson, D. (2007). Retest effects on identical and parallel forms

in certification and licensure testing. *Personnel Psychology*, 60, 367–396.
<https://doi.org/10.1111/j.1744-6570.2007.00077.x>

Raymond, M. R., Neustel, S., & Anderson, D. (2009). Same-form retest effects on credentialing examinations. *Educational Measurement: Issues and*

Practice, 28, 19–27. <https://doi.org/10.1111/j.1745-3992.2009.00144.x>

Wood, T. J. (2009). The effect of reused questions on repeat examinees. *Advances in Health Sciences Education*, 14, 465–473.
<https://doi.org/10.1007/s10459-008-9129-z>

Citation:

Babcock, B., & Siegel, Z.D.. (2022). Reinstatement Candidate Credentialing Exam Performance: Evaluating the Persistence of Misinformed Responses on Multiple Choice Items. *Practical Assessment, Research & Evaluation*, 27(8). Available online: <https://scholarworks.umass.edu/pare/vol27/iss1/8/>

Corresponding Author:

Ben Babcock
Elsevier

E-mail: b.babcock [at] elsevier.com