# The Time Crunch: How much time should candidates be given to take an exam?

**By Dr. Chris Beauchamp**                                                    **May 2016**

When developing an assessment, two of the major decisions a credentialing organization need to make are:
How many items will be on the exam, and
How much time will candidates be given to complete the exam.

These decisions can have a large impact on fairness and validity. Often, once an exam has been administered, many candidates will anecdotally report that they ran out of time and the assessment was unfair. So, an important question to ask is, what can credentialing organizations do in order to investigate and address these concerns?

In the credentialing field, assessments almost always have time limits. Although time limits are often set in order to provide standardized administration conditions, it is also important to allow candidates enough time to complete their assessment without undue time pressure. If the allocated time is too short, this could undermine validity (Lu & Sireci, 2007).

The purpose of this backgrounder is to explore this topic and provide a framework that credentialing organizations can use to determine if the time limits and length of the assessment promote fairness and validity.

Tests are classified as being either speed or power tests (Gulliksen, 1950). The difference between these two types of tests has impacts on what we call "test speededness", or the rate at which an exam is completed, as well as the correctness of candidate responses.

## Speed Tests:

A speed test is a test composed of items that are so easy that examinees would rarely give a wrong answer. However, the tests are so long that no candidate would complete the entire test in the allotted amount of time. As a result, candidates are judged by how far they get while taking the test before running out of time. This approach is commonly used in IQ Tests and other types of aptitude tests.

## Power Tests:

A pure power test is a test in which all items should be attempted, and the candidate's performance is judged by the correctness of their responses. Although most credentialing examinations are designed to be power tests, time limits are generally used. The question then becomes, "does the use of a time limit change a credentialing exam from being a power test to a speed test?"

Whenever tests involve a time limit, the rate at which candidates move through the items on the exam will affect their performance. A small percentage of candidates, no matter how much time is given, will not complete the exam. As a result, most examinations contain a mixture of speed and power components (Rindler, 1979).

There are several ways in which unintended test speededness may undermine test validity. For example, when speededness is unintended, candidate scores can be lowered due to factors such as anxiety and stress. Test speededness can also negatively impact content validity because some scored items are not attempted. This is problematic if the unattempted items fit into one or more content domains.

It is therefore important to demonstrate that a test is not overly affected by speededness. There are several statistical indices that are available to assess speededness using a single administration approach.

## Methods of assessing speededness

Educational Testing Service (ETS) provides three guidelines to assess test speededness (Schaeffer, Reese, Steffen, McKinley & Mills, 1993; Swineford, 1973). A test is speeded when 1) fewer than 80% of candidates do not complete the exam, 2) fewer than 100% of candidates reach 75% of the test and/or 3) the ratio of "not reached variance" to "total score variance" is greater than 0.15. The "not reached variance" represents the variance of the number of items left unanswered following the last item for which the candidate responded. This statistic is divided by the "total score" variance in order to obtain the "not reached" to "total score" variance ratio.

The three criteria presented above are based on the notion that a candidate will run out of time at the end of the exam and fail to respond to items that he or she has not reached. There are, however,

two other scenarios that should be considered. Near the end of the exam, a candidate may realize that there is insufficient time to finish the test. In one scenario, the candidate would accelerate their work rate and would skip items that would take too much time. This candidate would have a sporadic response pattern near the end of the exam. In another scenario, upon recognizing that time is running out, the candidate would respond to all the remaining items in a random manner (Oshima, 1994).

**Modification #1: Correcting for sporadic responding**

Under the original method, a candidate's stopping point is identified as when the candidate did not respond to any further items. However, it is possible that a candidate who is feeling significant time pressure will begin to answer questions sporadically. As a result, the original method should be modified. The candidate's stopping point can be identified as the point where the candidate responded to at least 3 questions in a row.

The following is an illustration based on a 200-item test:

| Question | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Response | 4 | 1 | 2 | 2 | | | | 1 | | | 1 |

*Original method:* The candidate is considered to have finished the exam.

*Refined method:* The candidate reached question 193 and did not complete the exam.

**Modification #2: Correcting for random responding**

Most candidates see a drop in performance near the end of the exam that is most likely due to fatigue. However, a candidate who is under significant time pressure may choose to either 1) respond to questions very quickly without fully reading or comprehending the material or 2) answer in a completely random manner. In some cases, a candidate may respond to every single question on the exam but may have responded randomly near the end.

To account for this, candidate performance on the first 25 and the last 25 scored items of the exam can be compared. Candidates who display a statistically significant ($p < 0.01$) drop in performance during the last 25 questions of the exam can be deemed to have run out of time. This is in contrast to the original method that would have considered this candidate to have finished the exam.

## What can be done to prevent an exam from being overly speeded?

Decisions related to the number of items on an exam and the time allocated to candidates to complete the exam should not be taken lightly. Significant research should be conducted first. This can include looking at historical response patterns for the exam or looking at comparable assessments. It would also be helpful, in the case of an entry-to-practice exam, to look at the policies in place at the educational/training level.

Some exams contain experimental/pilot items that are presented to candidates but do not count toward their total score. These items are typically presented throughout the exam. One strategy, if speededness is a concern, is to place these experimental/pilot items at the end of the exam (without indicating to candidates that the items are experimental/pilot items). This way, if candidates run out of time, this will not impact their exam scores because the unattempted items would be experimental/pilot items.

Many exam blueprints provide guidelines on the number of items from different competency categories. Items can then be presented in two ways: 1) items can be presented by competency category, or 2) items from different competency categories can be interspersed. If speededness were a concern, it would be important that the latter approach be used. This way, if candidates are unable to complete an exam, content validity impacts can be mitigated because items do not systematically come from one specific area.

In addition, many exam blueprints contain targets for taxonomy levels such as knowledge, application and critical thinking. In general, knowledge questions take less time to respond to compared to critical thinking questions. Therefore, taxonomy should also be considered when setting the blueprint.

Finally, from an exam administration perspective, candidates should be given frequent timing updates throughout the exam (e.g., "one hour left") and a clock should be visible. This will prevent candidates from losing track of their progress and running out of time.

## What if speededness is detected?

Evidence of speededness on an exam form may require the credentialing organization to revisit their exam blueprint to modify 1) the number of questions on the exam and/or 2) the time given to candidates to take the exam.

However, this does not address what to do with the candidates who took the exam that displayed evidence of excessive speededness. This is exacerbated in situations where numerous candidates are anecdotally reporting that the time limits were insufficient. What can be done to make the assessment fair for these candidates?

# Yardstick
## Testing and Training Experts

One solution would be to exclude the last handful of questions on the exam. For example, on a 200-item exam, the speededness analysis can be redone following the deletion of the last 10 questions. Does the revised 190-item exam still display evidence of speededness? If yes, it may be necessary to remove additional items.

## References

Gulliksen, H. (1950). Theory of mental tests. New York: John Wiley.

Kolen, M.J., & Brennan, R.L. (2004). Test equating, scaling, and linking. Methods and practices (2ed.). New York: Springer-Verlag.

Oshima, T.C. (1994). The effect of speededness on parameter estimation in item response theory. Journal of Educational Measurement, 31 (3), 200-219.

Rindler, S.E. (1979). Pitfalls in assessing test speededness. Journal of Educational Measurement, 16 (4), 261-270.

Schaeffer, G.A., Reese, C.M., Steffen, M. McKinley, R.L., & Mills, C.N. (1993). Field test of a computer-based GRE General Test (ETS Report No. RR-93-07). Princeton, NJ: Educational Testing Service.

Stafford, R.E. (1971). The speed quotient: A new descriptive statistic for tests. Journal of Educational Measurement, 8, 275-278.

**To read more articles related to eLearning, examination, and instructional design go to www.getyardstick.com and check out our blog.**