PSI eBrief #2
Number of Items on Exam Forms for multiple Choice questions Using Classical Test Theory

www.psionline.com

November 2014

Disclaimer. These guidelines are generalizations that are supported with evidence and do not represent an exhaustive list of supporting literature. Credentialing organizations should consult with a measurement professional since unique situations may require other acceptable evidenced based approaches and/or industry best practices. PSI believes this is a working document and any feedback is highly encouraged. Please send your feedback to **marketresearch@psionline.com**.

Factors to Consider	Summary	Supporting Literature
Internal Reliability	To meet minimum reliability requirements, the number of Items on a form should often be **21 or greater**. There are exceptions to this rule of thumb based on the breadth of the content domains (e.g., can a certain number of items on the exam form measure those domains adequately). Please note that quality of an item is a very important factor as well.	Cortina, J. M. (1993). What is Coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104.
Items to Be Developed (Item Banking)	"For selected response, a rule of thumb is that the item bank should **be 2.5 times** the size of a test."	Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge. Page 17
Testing Time	"Clear majority of examinees should have reached and attempted **90% or more** of the items in a test." Characteristics of the testing sample also plays a factor in the length of exam (e.g., items in German require more reading time than most other languages).	Schmeiser, C. B., & Welch, C. J. (2006). Test development. In Brennan, R.L. (Ed.), *Educational measurement (4th ed.)*. Westport, CT: Praeger. – Page 338

Distribution of Cognitive Items (optional)	"**Should be based on empirical data collected in a systematic way**" – such as that collected from the results of a Job Task Analysis/Practice Analysis.	Schmeiser, C. B., & Welch, C. J. (2006). Test development. In Brennan, R.L. (Ed.), *Educational measurement (4th ed.)*. Westport, CT: Praeger. – Page 316
Content	"In many cases the test domain must be prioritized to measure knowledge and skills judged to be most important by the relevant test audiences. The emphasis gathered through empirical survey data can serve as the basis for distributing items across these domains." The above quote by the authors can include other judgments/ratings other than importance such as frequency and criticality particularly for licensure exams (see Raymond, 2001).	1) Schmeiser, C. B., & Welch, C. J. (2006). Test development. In Brennan, R.L. (Ed.), *Educational measurement (4th ed.)*. Westport, CT: Praeger. – Page 319 2) Raymond, M. (2001). Job analysis and the specification of content for licensure and certification examinations. *Applied Measurement in Education, 14*(4), 369 – 415.

*These five factors should be considered simultaneously instead of independently.

ABOUT PSI

PSI has over 70 years of experience providing worldwide testing solutions to corporations, federal and state government agencies, professional associations, certifying bodies and leading academic institutions. PSI offers a comprehensive solutions approach from test development to delivery to results processing, including pre-hire employment selection, managerial assessments, licensing and certification tests, distance learning testing, license management services and professional services. More information is available at www.psionline.com.

800.367.1565