



**International
Credentialing
Associates**

**ICA eBRIEF #4
ITEM STATISTICAL FLAG GUIDANCE FOR
SELECTED RESPONSE QUESTIONS USING
CLASSICAL TEST THEORY**

www.intlcred.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. ICA believes this is a working document and any feedback is highly encouraged. Please send your feedback to manny@intlcred.com.

Factors to Consider	Guideline	Supporting Literature
Minimum number of item exposures / takes required	1) 50	1) Nunnally, J. C., & Bernstein, I. H. (1994). <i>Psychometric Theory</i> . New York: McGraw Hill. Page 301
	2) 50	2) Jones, Smith, & Talley (2006). Small-scale achievement testing systems. In Downing, S.M. & Haladyna, T.M. (Eds), <i>Handbook of test development</i> . New York, NY: Routledge. Page 497
	3) 20	3) Chinn, R. N., Hertz, N. R., & Showers, B. A. (2002) <i>Building and Managing Small Examination Programs</i> – Page 6 http://www.clearhq.org/resources/RB_SmallExamPrgs.pdf
Item difficulty ($P+$)	1) Acceptable range is 0.40 and 0.80	1) Osterlind, S. J. (1998). <i>Constructing test items: Multiple-choice, constructed response, performance, and other formats</i> . Boston: Kluwer Academic Publishers.– Page 269
	2) Acceptable item is greater than 0.30	2) Schmeiser, C. B., & Welch, C. J. (2006). Test development. In Brennan, R. L. (Ed.), <i>Educational measurement</i> (4th ed.). Westport, CT: Praeger. – Page 338
Key Point Biserial (r_{pb})	1) Accept items that are 0.20 or above	1) Schmeiser, C. B., & Welch, C. J. (2006). Test development. In Brennan, R. L. (Ed.), <i>Educational measurement</i> (4th ed.). Westport, CT: Praeger. – Page 339
	2) Accept items that are 0.20 or above	2) Nunnally, J. C., & Bernstein, I. H. (1994). <i>Psychometric theory</i> . New York: McGraw Hill –Page 396
	3) Accept items that are 0.15 or above	3) Haladyna, T. M., & Rodriguez, M. C. (2013). <i>Developing and validating test items</i> . New York, NY: Routledge. Page 350
	4) Reject items negative point-biserials	4) Osterlind, S. J. (1998). <i>Constructing test items: Multiple-choice, constructed response, performance, and other formats</i> . Boston: Kluwer Academic Publishers.– Page 279

Distractors	<p>1) Consider removing low response distractors. “Distractors that are seldom chosen by a test taker are useless and should be removed or replaced. These kinds of distractors are likely to be so implausible to all test takers that hardly anyone would choose them.” (page 352)</p> <p>2) If any option is selected 5% or less of the time it should be eliminated (page 354)</p> <p>3) “if distractors are implausible, they will not serve a useful function in measurement, but will usually be easily avoided by examinees.” (page 148)</p>	<p>1) & 2) Haladyna, T. M., & Rodriguez, M. C. (2013). <i>Developing and validating test items</i>. New York, NY: Routledge. Page 352, 354)</p> <p>3) Osterlind, S. J. (1998). <i>Constructing Test Items: Multiple-Choice, Constructed Response, Performance, and Other Formats</i>. Boston: Kluwer Academic Publishers.— Page 148</p>
-------------	---	---

Thank you. ICA would like to thank the individuals who provided feedback on this document, especially ICA’s Technical Advisory Board (www.intlcred.com/advisory.html).