Center for Advanced Studies in Measurement and Assessment

CASMA Technical Note

Number 7

A Note on Sample Size for Alternative Random Groups Equating Designs [1]

Michael J. Kolen

January 2015

[1]Thanks to Robert L. Brennan for his comments on an earlier draft of this note.

Center for Advanced Studies in
Measurement and Assessment (CASMA)
College of Education
University of Iowa
Iowa City, IA 52242
Tel: 319-335-5439
Web: www.education.uiowa.edu/casma

Contents

In random groups equating, often a set of new forms is spiraled with an old (anchor) form. However, having a large number of forms administered in any given test center can create test security issues. To address these security issues, it may be desirable to spiral subsets of forms in sets of test centers. However, doing so can lead to less equating precision for a given total sample size. This note is intended to provide a statistical framework for addressing this issue.

1 Method

Define

N as per form sample size for equating when all forms are spiraled, and

VAR as the equating error variance for equating scores on one form to scores on another form for a particular score.

Consider the following random groups equating design, where 7 forms are spiraled and Form A is the anchor (old) form:

Design 1
ABCDEFG

Assume that all forms are developed to the same content and statistical specifications. In this case, with sample size per form N, $VAR = VAR^*$ is the error variance for any of the equatings. The total sample size is $7N$, since there are 7 forms.

Consider an alternative equating design as follows:

Design 2
ABC
ADE
AFG

In this design, forms are spiraled within 3 separate sets of test centers, with Form A always being the anchor (old) form. In this case, again, the sample size per form is N and VAR^* represents the error variance for any of the equatings. However, because Form A appears two additional times, the total sample size is $9N$. Thus, the overall sample size needs are greater.

Consider another alternative as follows:

Design 3
ABC
BDE
DFG

In this design, the anchor (old) form for the first set of test centers is Form A, the anchor (old) form for the second set of test centers is B, and the anchor (old) form for the third set of test centers is D. Assume that the sample size for

each form in Design 3 is to be equal to a constant, N, with $VAR = VAR^*$ for any equating to the anchor (old) form in a row above.

Consider error variance in equating scores on Form F to scores on Form A. This equating involves a chain of the following 3 equatings: scores on Form F to Form D, scores on Form D to Form B, and scores on Form B to Form A. Generalizing from reasoning provided in Kolen and Brennan (2014, pp. 269-270), when the random groups design is used for equating, the equating error variance in equating scores on Form F to scores on Form A through a chain of 3 equatings is $VAR = 3VAR^*$ using a sample size of N per form. To obtain an error variance of this chain equal to VAR^*, the sample size would need to be multiplied by 3. That is, in this case the sample size for each equating needs to be $3N$, which leads to an overall sample size of $27N$, since there are 9 forms listed in the design.

Note that with a sample size of $27N$, the error variance for some equatings will be much less than VAR^* (for example, the equating of Form B to Form A would have an error variance of $VAR = VAR^*/3$), but any of the equatings to Form A will have $VAR \leq VAR^*$. Also, note that if the requirement stated above of having a constant N per form were relaxed, then it might be possible for the overall sample size to be substantially less than $27N$.

2 Example

Suppose that in Design 1, a sample size of 3,000 per form leads to an equating error variance of VAR^*. The overall sample size needed to equate all of the forms is 21,000 (7 x 3,000). The overall sample size for Design 2 needed to equate all of the forms with the same precision is 27,000 (9 x 3,000). The overall sample size for Design 3 to equate all of the forms with $VAR \leq VAR^*$ is 81,000 (27 x 3,000). This illustration demonstrates that (a) using one set of test centers is most efficient, and (b) if multiple sets of test centers are to be used in random groups equating, it is much more efficient to use the same anchor (old) form in each set of centers than to use different anchor (old) forms in each set of centers.

3 Reference

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking. Methods and practices.* New York: Springer.