

A Literature Review of Context Effect on Item Parameter Invariance (MARCES Team)

Topics	Findings (e.g., issues, concerns, and conclusions)
<ul style="list-style-type: none"> Definition of context effects (Yen, 1980) 	<ul style="list-style-type: none"> According to Yen (1980), if item parameters are influenced by item location, order effects, or the characteristics of other items in the test, there is evidence of <i>context effects</i>.
<ul style="list-style-type: none"> Factors investigated in context effects research 	<ul style="list-style-type: none"> Most research focuses on how context affects the stability of item parameters associated with item response theory (under the changing needs of the testing industry). For example, test disclosure and development of individual adaptive testing instruments have made the stability of item parameters a big concern especially with respect to equating.
<ul style="list-style-type: none"> Research in context effects shows inconsistent results with respect to item parameter estimates (Kingston & Dorans , 1984; Meyers, Murphy, Goodman & Turhan 2012; Rizavi, Way, Davey & Herbert, 2004; Store, 2013; Whitely & Dawis, 1976; Yen, 1980) 	<ul style="list-style-type: none"> Context effects affect item parameter estimates <ul style="list-style-type: none"> Item parameters estimated more than once from the same context are more highly related than those estimated from different contexts. Changing item positions between administrations negatively affects the measurement properties of a test. When item position changes substantially, some percentage of students can be classified into the wrong achievement level, which implies a big risk. Changes in item arrangements decreased the stability of item difficulty for both classical item difficulties (<i>p</i>-values) and Rasch item difficulty. The changes in item difficulty become especially salient for low ability examinees in comparison to high ability examinees. Context effects played a more significant role in item parameter estimates from computer adaptive tests when comparisons were made to the parameters that were obtained from paper-and-pencil testing. The linearly administered items exhibited remarkably small variation in parameter estimates over repeated calibrations.

<p>(Meyers, Kong & McClarty, 2008; Rubin & Mott, 1984)</p>	<ul style="list-style-type: none"> • Context effects do not always show significant effect on item parameter estimates <ul style="list-style-type: none"> • Item position change did not impact item difficulty or item discrimination.
<ul style="list-style-type: none"> • Context effects with respect to equating <p>(He, Gao, and Ruan, 2009; Kingston and Dorans, 1984; Meyers, Murphy, Goodman & Turhan 2012)</p>	<ul style="list-style-type: none"> • Equating results were different when including some item types that are more sensitive to location changes. • Any differences among the equating results were due to the way the items were pre-tested, contextual/order effects, or violations of IRT assumptions. • The effort to keep the items used for equating in identical or very similar positions within the test from one administration to the next can be difficult due to test security concerns, test disclosure requirements, limitations in the item bank, or test content and design considerations. • Large changes in item positions between administrations can have great impact on both item parameter estimates and the subsequent equating results.
<ul style="list-style-type: none"> • Summary of findings 	<ul style="list-style-type: none"> • Whenever items exhibit within-test context effects, they should maintain the same location on the new form as they had in the old form. • Item types sensitive to location changes will be inappropriate for use in an adaptive testing context. • Test constructors continue the general policy of building tests with items most recently field tested, but they might also consider using up to 3-year-old items when those items meet their needs better than more recent items (Based on research done on context effects on field-testing and operational testing). • Test developers should try to keep items in similar or as close to the original positions across administrations since when item position changes substantially, some percentage of students can be classified into the wrong achievement level, which can be a big risk.

References

- Meyers, J. L., Kong, X. J., & McClarty, K. L. (2008). *An investigation of changes in item parameter estimates for items re-field tested*. Paper presented at the Annual Meeting of the American Educational Research Association, New York.
- Meyers, J. L., Murphy, S., Goodman, J., & Turhan, A. (2012). The impact of item position change on item parameters and common equating results under the 3PL model. Presented at the Annual Meeting of the National Council on Measurement in Education, Vancouver, B. C.
- Store, D. (2013). *Item parameter changes and equating: An examination of the effects of lack of item parameter invariance on equating and score accuracy for different proficiency levels*. Unpublished doctoral dissertation, the University of North Carolina.
- Whitely, E., & Dawis, R. (1976). The influence of test context on item difficulty. *Educational and Psychological Measurement*, 36, 329–337.
- Yen, W. M. (1980). The extent, causes and importance of context effects on item parameters for two latent trait models. *Journal of Educational Measurement*, 17, 297-311.