Passage Effects on Curriculum-Based Measures in Reading

Esther H. Yi, B. A. and Kara M. Styck, Ph.D.

Illinois School Psychology Association Convention, January 30, 2020, Springfield IL

Department of Psychology, Northern Illinois University

Northern Illinois University

Introduction

Grades K-3 presents a unique window of opportunity wherein reading difficulties are most receptive to instruction (Simmons et al., 2008). Results of intervention research suggests that children who start out as poor readers in the beginning of kindergarten respond quickly and positively to early intervention. These young struggling readers demonstrate positive, generalized reading outcomes when given intensive reading interventions (Wanzek et al., 2018). Continuous progress monitoring plays a critical role in multi-tiered systems of supports (MTSS) that have emerged as a way to ensure high-quality universal reading instruction for all students and to provide differentiated levels of instructional support to struggling readers who need early intervention (Greenwood et al., 2014).

Progress monitoring academic performance is primarily conducted using curriculum-based measures (CBM; Deno, 1985). Oral reading fluency (ORF; also referred as R-CBM) passages are the most commonly used CBM to monitor student progress in reading (Fuchs et al., 2001). R-CBM are a standardized set of individually administered passages intended to assess word reading accuracy and fluency with connected text. The most commonly referenced score obtained from R-CBM is the number of words a student reads correctly in one minute (WCPM), which is strongly related to general reading proficiency (Fuchs et al., 2001). There is also a growing body of evidence demonstrating that WCPM is a better measure of specific reading subskills (e.g., reading comprehension, phonemic awareness, etc…) than specific subskill mastery measures themselves (Ardoin et al., 2004; Fuchs et al., 2001; Van Norman et al., 2018). However, Fuchs et al. (2001) cautioned that the utility of WCPM as an indicator of overall reading competence is dependent upon the assumption of equal passage text difficulty of within-grade-level passages.

Passage Effects

Differences in passage difficulty are problematic because they imply that WCRM scores measure more than a student's ORF ability—they measure ORF ability and passage text difficulty. Specifically, 10-11% of variance in WCPM scores can be attributed to variability in passage text difficulty level (Chaparro et al., 2018; Poncy et al., 2005). This error variance due to differences in text difficulty levels has been referred to by Cummings, Park, and Bauer Schaper (2013) as passage effects. Passage effects undermine the validity of any educational decisions made based on WCPM (Albano & Rodriguez, 2012). This means that two students with the same latent ORF ability who are administered R-CBM passages with varying passage difficulty levels may achieve different WCPM scores simply because one student was administered a relatively easier/difficult passage than the other student.

Examples

Screening

Chaparro et al. (2018), for example, reported that two second-grade students reading at the same grade level who were administered different passages obtained scores that were up to 22 WCPM apart because they were administered passages with different text difficulty levels.

Student	ORF Ability	Passage	WCPM
Arya	90	A	75
Dale	90	B	95

Passage effects also impact the measurement precision of individual students' WCPM scores. Poncy et al. (2005) estimated the standard error of measurement (SEM) for mean WCPM scores of DIBELS R-CBM third grade passages for 37 third-grade students and discovered that the least precise WCPM scores were obtained when only one passage was administered (SEM 18 WCPM) and the most precise WCPM scores were obtained when nine passages were administered (SEM 6 WCPM).

Student	Number of Passages	WCPM	95% Confidence Interval
Arya	1	60	26-97
Arya	9	60	50-73

Francis et al. (2008) examined the impact of passage effects in another commonly suggested CBM practice of taking the median WCPM score from three passages with a sample of 134 second-grade students. They reported that, in a set of 20 progress-monitoring passages, none were equally as likely to produce a student's median WCPM score. In fact, some passages were more likely to be selected as the median passage than others, especially when they were paired with an easier passage and a more difficult passage.

Progress Monitoring

Additionally, Francis et al. (2008) reported that students who were administered the most difficult passage exemplified a decreasing trend in scores; whereas, students who were administered the easiest passage exemplified an increasing trend in scores. True rates of growth in ORF, therefore, were masked when passages varied in difficulty level. In short, administering a relatively easier/difficult passage at one progress-monitoring time point than the previous progress-monitoring time point may erroneously cause a student's reading rate to improve, decline, or stay the same, irrespective of whether or not learning has taken place (Ardoin & Christ, 2009; Francis et al., 2008).

Addressing Passage Effects

Readability Formulas

Readability formulas are used as an a priori equating method, meaning that passages are either written or assigned to grade levels according to their identified readability level (Cummings et al., 2013). These formulas are relatively simple and easy to use, but have been largely criticized as an inadequate means to equate R-CBM passage scores because the same passage may have a different readability level depending on what formula is applied.

Statistical Equating

Statistical equating using classical test theory (CTT) adjusts WCPM on one (or more) passages to the scale of referent values. Referent values can be derived by matching the test-score distributions of comparison passages to that of a referent passage, or the average of all passage scores at a grade, in terms of their mean, standard deviation, and/or percentiles. Accordingly, these procedures can take one of three forms: (a) mean equating, (b) linear equating, or (c) equipercentile equating. Theoretical assumptions, sample size requirements, and amount of measurement error guide decisions to select the best equating method based on CTT to use. It otherwise poses serious limitations.

Statistical equating using item-response theory (IRT), on the other hand, measures a person's latent trait or ability as the probability of the particular person answering a particular test item correctly (von Davier, 2011). IRT models rely on separate person and item parameters, such that they are estimated independent of who takes the test or which test is given. Within this framework, students with the same ORF ability, for example, will have the same probability of reading a passage correctly, regardless of other students' abilities or other passage difficulty levels. Using an IRT-based equating method may overcome the limitations posed by CTT-based equating methods.

Conclusions

1) Passage effects exist within and across grade-level passages. This means that passages within and across grade levels very likely meaningfully differ in difficulty level in a way that confounds measurement of student oral reading fluency (ORF) ability.

2) Decisions based on inaccurate WCPM scores can result in withholding additional reading supports for students who truly need them and incorrectly identifying students as needing more intensive supports that are unnecessary, thereby taxing already limited need-based resources (e.g., time, staff, intervention program costs).

3) Administering more passages in one administration and/or taking the median WCPM as the final score does not ameliorate passage effects. In fact, Francis et al. (2008) claimed that selecting a median passage essentially discards the data from the other two passages in favor of one single R-CBM passage and negates any benefit of administering more than one passage.

4) Two primary methods have been used to equate WCPM on passage text difficulty: (a) readability formulas and (b) statistical equating procedures. Readability formulas have been widely criticized as not significantly related to a student's WCPM and as an inadequate equating method.

5) Statistical equating using classical test theory (CTT) accounts for passage effects by adjusting WCPM on one (or more) passages to the scale of referent values. It is recommended that the same referent values are used when using CTT-based statistical equating, whether it be a referent passage or the average of all passage scores for a grade level.

6) Statistical equating using item-response theory (IRT) overcomes limitations posed by CTT-based equating methods. However, it is yet to be widely researched in the context of equating WCPM scores.