Encyclopedia of Information Science and Technology, Fourth Edition

Mehdi Khosrow-Pour
Information Resources Management Association, USA

IGI Global
DISSEMINATOR OF KNOWLEDGE

Published in the United States of America by
IGI Global
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Names: Khosrow-Pour, Mehdi, 1951- editor.
Title: Encyclopedia of information science and technology / Mehdi Khosrow-Pour, editor.
Description: Fourth edition. | Hershey, PA : Information Science Reference, [2018] | Includes bibliographical references and index.
Identifiers: LCCN 2017000834| ISBN 9781522522553 (set : hardcover) | ISBN 9781522522560 (ebook)
Subjects: LCSH: Information science--Encyclopedias. | Information technology--Encyclopedias.
Classification: LCC Z1006 .E566 2018 | DDC 020.3--dc23 LC record available at https://lccn.loc.gov/2017000834

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

The Relationship Between Online Formative Assessment and State Test Scores Using Multilevel Modeling

Aryn C. Karpinski
Kent State University, USA

Jerome V. D'Agostino
The Ohio State University, USA

Anne-Evan K. Williams
Billings Middle School, USA

Sue Ann Highland
Grand Canyon University, USA

Jennifer A. Mellott
Kent State University, USA

INTRODUCTION

The main goal of the current study was to examine the relationship between online formative assessments (FAs) and summative, yearly state proficiency test scores. Specifically, the relationship between one online formative assessment (FA) program in reading, known as the Diagnostic Online Reading Assessment (DORA), and state test scores in reading (i.e., the Colorado Student Assessment Program [CSAP]) was examined in four cohorts across elementary, middle, and high school in beginning in the 2004/2005 academic year and ending in 2009/2010. This investigation used Hierarchical Linear Growth Modeling (HLGM; i.e., Multilevel Modeling) to address the following research question: (1) What is the relationship between online formative assessment score growth and state test score growth?

Formal and informal FAs are one of many teaching methods that have been used to increase student performance on end-of-course, academic year, and other high-stakes achievement tests for decades and has a large research base to support these practices (e.g., Black & Wiliam, 1998a). Additionally, summative assessment data (e.g., yearly state proficiency tests) are continually used as indicators of school and district performance for policymakers and the public. However, these summative data are of little use in the day-to-day activities of teachers in diagnosing student learning progress and modifying teaching strategies, as is done in the FA process (e.g., Black, 2015). Because this collection of abstract theories and research methods have transitioned into actual teaching practices, it is important to build the literature surrounding technology-based methods as teachers continue to use FA in the classroom.

The purposes in conducting this study include the following: (1) To support the burgeoning literature outlining the role of technology in general in teaching and learning, and (2) To bolster support for federal initiatives and administrative demands for more efficient ways to meet state standards. As technology-based assessment is gradually used to support and/or replace traditional forms of evaluation, the need to examine the extent to which these methods are educationally sound is

DOI: 10.4018/978-1-5225-2255-3.ch450

in high demand. Overall, information presented in this study can provide practical implications to district-wide implementation of supplemental reading instruction in an online environment.

BACKGROUND

E-learning (i.e., learning that is facilitated by electronic technologies) is referred to as part of the equipment of 21st Century scholarship (Buzzetto-More & Guy, 2006). However, e-learning is only half of the equation as government mandates have required schools to use data to inform decision making. The use of data has necessitated the development of improved information technology and access to computers and high-speed Internet in schools (Petrides, 2006). Thus, the other half of the equation is the use of data rendered from e-learning, or e-assessment, which entails using electronic technologies to drive student learning and assessment as with FA (Ridgway, McCusker, & Pead, 2004).

FA can be briefly defined as the use of diagnostic formal and informal assessments to provide feedback to teachers and students over the course of instruction for the purpose of improving performance and achievement (e.g., Black, 2015; Boston, 2002). Previous research in this area has primarily focused on traditional FA practices (e.g., paper-and-pencil quizzes), with the current literature beginning to examine the effectiveness of Internet-based, automated FA programs (e.g., Chua & Don, 2013; Kingston & Nash, 2011). The overall consensus from the traditional body of literature is that FA is an essential component of classroom procedure, and that its proper use can raise standards and achievement (e.g., Black & Wiliam, 1998a; Carlson, Borman, & Robinson, 2011; Gulikers, Biemans, Wesselink, & van der Wel, 2013; Merino & Beckman, 2010), with the latest studies of technology-based FA beginning to echo these findings. Many theories have attempted to describe FA in terms of multilevel relationships (i.e., students, teachers, schools, school districts, etc.), with few studies focusing on statistically accounting for these nested associations, and hardly any examining technology-based FA practices (Black & Wiliam, 2009).

Regarding previous studies of online FA, the overwhelming majority of these studies have examined college-age populations in the university setting, usually within one course (e.g., Buchanan, 2000; Jenkins, 2004). In addition, past and current FA research has thoroughly examined the relationship between measures of FA and performance on a summative, usually end-of-course or final exam, but not state proficiency test scores. This area of research is just beginning to use more sophisticated statistical analyses, which is in contrast to the many qualitative studies summarizing student perceptions of a technology-based platform for quizzes/exams (Hunt, Hughes, & Rowe, 2002; Peat & Franklin, 2002). Additionally, due to the novelty of the mode of online or computerized administration, understandably research is lacking in longitudinal data analysis, with few studies examining multiple years of data across several cohorts.

METHODS

Existing DORA data were provided from one school district in Colorado from an online testing company, and existing CSAP data were provided from the same school district by the Colorado Department of Education. The data were selected via collaboration with the testing company and one school district in Colorado. The selected school district was one that gave permission to use their student demographic information and state test scores. It was necessary to have permission from both parties as the only way to examine correlated growth is to link the data via a shared student ID number. Additionally, this particular school district was selected because they had fewer missing data, with all students having graduated or left the school district at the present time. Data were linked anonymously producing four cohorts of students across grades 3 through 11. It was hypothesized that student FA score growth will be significantly and positively related to student state test score growth.

The analytical method used was a *Two-Level Time-Varying Covariate* HLGM (Singer & Willett, 2003). According to Raudenbush and Bryk (2002), investigating a relationship with a growth trajectory of another variable of interest is common practice in multilevel modeling. Time-varying covariates are defined as person-level characteristics that are measured and may change over time, and are related to the outcome (O'Connell & McCoach, 2004). Thus, the measurements across time and other Level 1 predictors form a nested structure when combined with other student level variables (i.e., Level 2).

Time (i.e., in months with one unit being every 3 to 6 months) and DORA scores (i.e., the time-varying covariate) were used in Level 1. Demographic covariates such as gender (i.e., SEX), ethnicity (i.e., ETHNIC), SES (i.e., FREERED, which is Free/Reduced lunch status), and ESL/ELL status were incorporated into Level 2 of the model. CSAP scores were used as the outcome variable (see Table 1). Four models were run using the DORA subtests (i.e., Word Recognition [WR], Oral Vocabulary [OV], Spelling [SP], and Reading Comprehension [RC]) as the time-varying covariate in separate models. These subtests were chosen because they did not have ceiling effects compared to the other subtests.

DATA

Existing DORA and CSAP data were used (N = 208) from 2004/2005 to 2009/2010 for students in grades 3 through 11. For the CSAP, students' test scores from the spring of 2005 were the first data point. Two state reading test scores were obtained before DORA implementation and three afterwards. For DORA, students' test scores from fall of 2006 represented the first data point.

Table 1. Hierarchical linear growth model level 1 and level 2 formulae and explanation of symbols

Formula Component	Explanation
Level 1 Model	$Y_{ti} = \pi_{0i} + \pi_{1i}(DORA)_{ti} + \pi_{2i}(Time)_{ti} + e_{ti}$
Y_{ti}	Student's CSAP score for time t for student i
$(Time)_{ti}$	Elapsed years/months since DORA implementation
$(DORA)_{ti}$	Time-varying predictor for a student at a given time point
π_{0i}	Intercept/Student's initial CSAP score
π_{1i}	Linear growth coefficient (for DORA)
π_{2i}	Growth rate over all years/months
e_{ti}	Individual student error
Level 2 Model	$\pi_{0i} = \beta_{00} + \beta_{01}(SEX)_i + \beta_{02}(ETHNIC)_i + \beta_{03}(ESLELL)_i + \beta_{04}(FREERED)_i + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(SEX)_i + \beta_{12}(ETHNIC)_i + \beta_{13}(ESLELL)_i + \beta_{14}(FREERED)_i + r_{1i}$ $\pi_{2i} = \beta_{20} + \beta_{21}(SEX)_i + \beta_{22}(ETHNIC)_i + \beta_{23}(ESLELL)_i + \beta_{24}(FREERED)_i + r_{2i}$
π_{0i}	Individual-specific CSAP score parameter (initial status)
π_{1i}	Individual-specific CSAP score parameter (DORA growth)
π_{2i}	Individual-specific CSAP score parameter (growth rate)
β_{00}	Baseline expectation (initial CSAP status) for the demographics coded 0
β_{10}	Expected change of the CSAP controlling for the DORA time-varying covariate
β_{20}	Expected change of the CSAP for the demographic predictors coded as 0
r_{0i}, r_{1i}, r_{2i}	Residuals

Students were tested approximately three times during the school year, with a possible total of 11 DORA assessments for the years investigated in the current study. See Table 2 for the sample demographic information.

MEASURES

DORA tests were across seven subtests: (1) High-Frequency Words, Phonics, Phonemic Awareness, Word Recognition (WR), Oral Vocabulary (OV), Spelling (SP), and Reading Comprehension (RC). With the exception of the SP subtest which is a generative test, all test items are multiple-choice. DORA results are returned as grade-level equivalencies. The CSAP is administered each spring, yielding a single, scaled score (i.e., reading score in the current study) for each student every year. The state scores were based on a scale that ranged from 0 to 1000 depending on the grade level assessed, with a score of approximately 550 as the cut-point for proficiency each year. At each grade level, students are assessed using 40 to 70 multiple choice items depending on the grade level, developed to assess student knowledge of grade-level indicators, identified as the Colorado Model Content Standards for that particular grade level. The tests across grades were vertically equated.

Table 2. Demographic information of the sample (N = 208) for grades 3 through 11 across the 2004/2005 to 2009/2010 academic years

Demographic Information	*n* (%)
Cohort	
1	47 (22.6)
2	52 (25.0)
3	48 (23.1)
4	61 (29.3)
Gender	
Male	104 (50.0)
Female	104 (50.0)
Ethnicity	
White (Non-Hispanic)	142 (68.3)
Hispanic	60 (28.8)
Black (Non-Hispanic)	3 (1.4)
Asian/Pacific Islander	2 (1.0)
American Indian/Alaskan Native	1 (.5)
Free/Reduced Lunch	
Eligible	93 (44.7)
Not Eligible	115 (55.3)
English as a Second Language/English Language Learner	
Yes	26 (12.5)
No	182 (87.5)

RESULTS

The main assumptions in Hierarchical Linear Growth Modeling pertain to the functional form of the model examining linearity, and the stochastic part of the model involving normality and homoscedasticity. Compound symmetry (i.e., a sufficient condition for sphericity; Maxwell & Delaney, 1990) can be relaxed in a multilevel framework (Raudenbush & Bryk, 2002). Linearity, normality, and homogeneity of variance (i.e., homoscedasticity), however, are assumptions that are typical of any General Linear Model (GLM) approach. Thus, these assumptions were checked and upheld to ensure unbiased estimates of population effects.

The full model-building strategy was implemented for each DORA subtest as the time-varying covariate (i.e., One-way Random Effects Analysis of Variance [ANOVA], Unconditional Model, Conditional Growth Model, Full Model; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Only the Full Models will be discussed below (see Table 3). Each column in the table provides the results from one HLGM analysis (i.e., one per DORA subtest). The first row in the table provides the estimated average student DORA scores at Time 0, the standard error for these estimates (in parentheses), and whether the estimates were significantly greater than zero.

Table 3. Hierarchical linear growth models for Colorado Student Assessment Program (CSAP) scores and time predicting online formative assessment (Diagnostic Online Reading Assessment [DORA]) subtests (N = 208)

Fixed Effects	Word Recognition	Oral Vocabulary	Spelling	Reading Comprehension
Initial CSAP Status (π_{0i})				
Intercept (β_{00})	634.38 (4.94)***	632.19 (5.08)***	637.47 (4.66)***	634.33 (4.70)***
Sex (β_{01})	5.94 (6.27)	9.66 (6.17)	8.40 (5.85)	5.33 (6.03)
Ethnicity (β_{02})	-2.85 (7.80)	4.75 (7.10)	-4.49 (7.17)	3.15 (7.32)
ESL/ELL (β_{03})	-36.56 (10.51)**	-36.58 (11.23)*	-41.19 (10.30)***	-20.37 (11.35)
Free/Reduced Lunch (β_{04})	-16.13 (6.65)*	-16.73 (6.44)**	-16.24 (6.13)**	-16.88 (6.79)*
DORA Growth Rate (π_{1i})				
Intercept (β_{10})	1.22 (.84)	.31 (.84)	4.16 (1.09)***	2.93 (.66)***
Sex (β_{11})	.78 (1.03)	2.77 (1.18)*	1.81 (1.33)	-.14 (.94)
Ethnicity (β_{12})	-1.05 (1.58)	2.99 (1.60)	-1.47 (1.57)	1.05 (1.13)
ESL/ELL (β_{13})	2.01 (1.97)	-1.66 (2.23)	-2.30 (2.31)	1.91 (1.65)
Free/Reduced Lunch (β_{14})	.44 (1.02)	.36 (1.36)	-.90 (1.45)	-.64 (1.02)
CSAP Growth Rate (π_{2i})				
Intercept (β_{20})	11.89 (1.45)***	12.43 (1.52)***	9.66 (1.52)***	9.15 (1.39)***
Sex (β_{21})	.78 (1.88)	-1.20 (2.03)	-.97 (1.99)	1.68 (1.97)
Ethnicity (β_{22})	.13 (2.25)	-2.97 (2.18)	-.43 (2.24)	-2.44 (2.25)
ESL/ELL (β_{23})	-1.18 (3.45)	1.91 (3.58)	2.69 (3.69)	-3.67 (3.98)
Free/Reduced Lunch (β_{24})	.61 (1.89)	1.09 (1.96)	2.48 (1.88)	2.42 (2.12)

Note. *p <.05; **p <.01; ***p <.001. The columns are the four DORA reading subtests. ESL/ELL = English as a Second Language/English Language Learner.

All initial DORA average scores across the four subtests were significantly greater than zero. The growth rates for all subtests were significantly greater than zero. Finally, the DORA and CSAP covariation results indicated that student gain in DORA over time covaried positively and significantly with their CSAP gain on the SP and RC subtests. In comparing the growth rates for SP and RC, both were significantly positively related to the state test in reading. For every one-point increase in SP score, there was a 4.16-point increase in the state test score, and for every one-point increase in RC score, there was a 2.93-point increase in the state test score. Therefore, student performance on the SP subtest resulted in faster growth on the CSAP compared to RC.

DISCUSSION

The goal in this study was to examine if student CSAP growth is related to student DORA growth. The hypothesis was that student CSAP growth will be significantly and positively related to student DORA growth. Thus, the relationship of interest in addressing this research question is with the time-varying covariate and the state reading test. The hypothesis was partially supported in the Full Models controlling for the demographics in that the DORA scores for Spelling and RC were positively and significantly related to state reading test scores, indicating that these subtests are demonstrating a correlated growth in students reading to the state testing regardless of demographic status.

As noted in the results section, the time-varying growth rates for SP and RC from the Full Models were both significantly and positively related to the state test in reading; however, student performance on the SP subtest had faster growth on the state test compared to RC. This is an interesting finding as typically the RC subtest is viewed as the most similar in structure and content to state reading tests (Let's Go Learn, Inc. ©, 2013). The RC subtest attempts to access the semantic domain of a learner's reading abilities. Children silently read passages of increasing difficulty and answer questions about each passage immediately after they read it. The questions for each passage are broken up into three factual questions, two inferential questions, and one contextual vocabulary question. This is typically how many state reading tests structure their exams.

As indicated above, the SP subtest surprisingly was related to the fastest state test score growth rates in students. SP is a generative process as opposed to a decoding or meaning-making process as seen in most assessments of reading comprehension, which does not support the finding as noted above. Additionally, it is natural for young readers' spelling abilities to lag a few months behind their reading comprehension abilities (Bear, Invernizzi, Templeton, & Johnston, 2000). Overall, the significant findings from the Full Models indicate that as modes of online FA, SP and RC online FA subtests are related to faster state test score growth rates, with SP producing the highest growth rate.

With regards to the non-significant findings for WR and OV, it is not alarming to see that the WR subtest growth was not correlated with CSAP growth, as the testing of word identification skills out of context is typically not a skill that is the focus of standardized reading assessments (Let's Go Learn, Inc. ©, 2013). As for OV, a significant correlation between this subtest and state test score growth was expected, as a learner's knowledge of words and what they mean is an important part of the reading process (Butler, Marsh, Sheppard, & Sheppard, 1985). The knowledge of word meanings affects the extent to which the learner comprehends what is read, such as in more traditional standardized reading tests. This subtest asks students to select the picture that correctly corresponds to a word they hear. This audio-visual format may explain the non-significant relationship between the more standardized paper-and-pencil format for most reading state tests.

SOLUTIONS AND RECOMMENDATIONS

Although causal inference is limited, the demonstrated relationship can provide teachers and administrators with evidence to warrant the continued use of technology-based FA practices. Specific to the results, the SP and RC subtests are further supported as a learning tool to gauge, or perhaps predict, student performance on the reading state test. The fact that DORA use was significantly and positively related to the state test even when student demographics were controlled, suggests that DORA SP and RC have something additional to offer.

For teachers/educators, focusing on a student's SP and RC growth can potentially add to growth on the state reading test. For example, if a teacher can raise students' DORA SP subtest score by one point, he or she can expect to see a 4.16-point increase on the reading state test on average. And if a teacher can raise students' DORA RC subtest score by one point, he or she can expect to have a 2.93-point increase on the state test (i.e., every three to six months). Thus, if a state reading test is given once a year (i.e., every 12 months) and this mode of online FA renders a one-point increase every three to six months, students' scores on the state test are predicted to grow between 8.32 and 16.64 points that year.

Teachers (and administrators) will also benefit from the results by garnering support for the use of online FA from a practical perspective. One major benefit is the ease of disseminating feedback to students after an assessment, and using the automated, specialized feedback to diagnose problems and quickly remedy these issues in time for the state exam. Buchanan (2000) noted that the individualized feedback makes this mode of online FA ideal for large, multi-section, introductory-level college courses. In the case of the current study, this mode can also be deemed ideal for large classrooms of elementary, middle, and high school students, in which their teachers may not have the time or resources to give specialized feedback to everyone.

Another implication is the practical advantage of ease of assessment for the large number of students being tested in the educational system. For FA to be most effective, quality feedback should be provided at frequent intervals, and testing a large number of students frequently with specialized feedback can advocate the use of a technology-based mode of FA. Implications also extend to the cost surrounding mass testing. Since a positive relationship was indicated between online FA scores and state test scores, this may allow administrators to have the necessary support to purchase site licenses and invest money in such programs, which are generally cheaper to administer frequently in bulk.

For administrators, the demand for school systems, individual schools, and teachers to be accountable for student performance has increased considerably over the past two decades. This demand for accountability relates to a direct measurement of attainment of educational standards and objectives. The results from this research question support the use of online FA tools as a way to measure and attain various educational standards such as having students pass and excel on the end-of-year, summative state exam. Overall, these results provide some support for administrative demands to find more efficient ways to meet state standards. The significant and positive relationship between the scores may help get support for schools to obtain the funds needed for programs to alleviate some constraints of mass assessment.

Society, in general, benefits from these findings by demonstrating how online FA practices may have the potential to give educators a more efficient and consistent way to make accurate, data-driven, informed decisions when evaluating student progress. This technology-enhanced efficiency has the potential to support independent student learning and facilitate future lifelong learners in our society. Rarely do teacher practices and methods have such depth of empirical support and is considered a universal best practice. Because of this widespread acceptance, evidencing how technology-based methods of FA influence achievement becomes paramount in demonstrating student learning gains and success in school and beyond.

FUTURE RESEARCH DIRECTIONS AND LIMITATIONS

Although the usefulness of correlational studies and related research questions have a place in the research process, causal conclusions cannot be stated. Future research should consider implementing a similar design, but also obtain an adequate control group. With regards to general threats to internal validity, some limitations are apparent (Cook & Campbell, 1979). Maturation is a limitation, since the existing data sampled took place over months in academic years (Kazdin, 2003). Another methodological limitation includes the use of one school district. Future studies should include multiple school districts from a range of rural and urban areas and involve public and private schools as well.

With regards to the DORA and CSAP data, the DORA data are from 2006/2007 to current and the CSAP data range from 2004/2005 to present as well. Moreover, there are more DORA time points than CSAP time points. Thus, data were not collected at the same time points (i.e., it was approximated). Generally, HLGMs can accommodate time-unstructured data such as the above; however, the accuracy and validity of results may depend on how closely the data are measured (i.e., same time/day compared to several days/weeks apart) in a time-varying covariate model (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). Although this could be considered problematic, analyzing only the data points that "matched" compared to all the data did not change the substantive results.

Analyzing only students in grades 3 through 11 may be considered another weakness. Students in grades Preschool through 2 and grade 12 were not included because this study focused on the state test and regularly administered FAs, which only occur between grades 3 through 10. State testing in Colorado begins in grade 3, and grades 11 and 12 are given college preparatory exams and high school exit exams. Additionally, DORA is administered more frequently in younger grade levels, and at least three time points are necessary to analyze the data. Future studies should consider analyzing all grade levels with complete data from multiple districts.

CONCLUSION

The body of FA literature has unanimously heralded the benefits of the diagnostic use of assessment to inform curriculum and instruction, and consequently, improve student performance and achievement. Previous research in this area has primarily focused on traditional FA practices. More recently with the technology movement in schools, the literature is beginning to examine the effectiveness of Internet-based FA, with the latest studies of this modern mode FA beginning to replicate these findings. The current study attempted to add to this literature base by examining one online FA program and its relationship to a summative state proficiency test.

It was hypothesized that student online FA growth would be related to state test score growth. This hypothesis was partially supported in that these online FA scores demonstrated a correlated growth with the state test scores regardless of demographic status. The findings can provide some support to the burgeoning literature outlining the role of online FA in teaching and learning. Internet-mediated teaching and assessment is becoming commonplace in the classroom, and is more frequently being used to replace traditional modes of student assessment. The need to examine the extent to which these methods are educationally sound is in high demand. Results from this study can not only add to the literature base theoretically and methodologically, but also practically, by bolstering support for federal initiatives and administrative demands for more efficient, technology-based ways to encourage teachers to invest their time in this mode of FA, and in turn, meet state standards and increase student achievement.

REFERENCES

Bear, D. R., Invernizzi, M., Templeton, S., & Johnston, F. (2000). *Words their way: Word study for phonics, vocabulary, and spelling instruction* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*(1), 30–52. doi:10.1037/1082-989X.9.1.30 PMID:15053718

Black, P. (2015). Formative assessment – An optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 161–177. doi:10.1080/0969594X.2014.999643

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–75. doi:10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. doi:10.1007/s11092-008-9068-5

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation*, *8*(9). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=9

Buchanan, T. (2000). The efficacy of a world-wide web mediated formative assessment. *Journal of Computer Assisted Learning*, *16*(3), 193–200. doi:10.1046/j.1365-2729.2000.00132.x

Butler, S., Marsh, H. W., Sheppard, M. J., & Sheppard, J. L. (1985). Seven-year longitudinal study of the early prediction of reading achievement. *Journal of Educational Psychology*, *77*(3), 349–361. doi:10.1037/0022-0663.77.3.349

Buzzetto-More, N., & Guy, R. (2006). Incorporating the hybrid learning model into minority education at a historically black university. *Journal of Information Technology Education*, *5*, 153–164.

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378–398. doi:10.3102/0162373711412765

Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers motivation. *Computers in Human Behavior*, *29*(5), 1889–1895. doi:10.1016/j.chb.2013.03.008

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Gulikers, J. T. M., Biemans, H. J. A., Wesselink, R., & van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation*, *39*(2), 116–124. doi:10.1016/j.stueduc.2013.03.001

Hunt, N., Hughes, J., & Rowe, G. (2002). Formative automated computer testing (FACT). *British Journal of Educational Technology*, *33*(5), 525–535. doi:10.1111/1467-8535.00289

Jenkins, M. (2004). Unfulfilled promise: Formative assessment using computer-aided assessment. *Learning and Teaching in Higher Education*, *1*, 67–80.

Kazdin, A. E. (2003). *Research design in clinical psychology*. Boston, MA: Allyn & Bacon.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*(4), 28–37. doi:10.1111/j.1745-3992.2011.00220.x

Let's Go Learn, Inc. (2013). *Let's Go Learn Research.* Retrieved November 21, 2013, from http://www.letsgolearn.com/lglsite/research

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

Merino, K., & Beckman, T. (2010). Using reading curriculum-based measurements as predictors for the measure academic progress map standardized test in Nebraska. *International Journal of Psychology: A Biopsychosocial Approach, 6*, 85 – 98.

OConnell, A. A., & McCoach, D. B. (2004). Applications of hierarchical linear models for evaluations of health interventions: Demystifying the methods and interpretation of multilevel models. *Evaluation & the Health Professions*, *27*(2), 119–151. doi:10.1177/0163278704264049 PMID:15140291

Peat, M., & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology*, *33*(5), 515–523. doi:10.1111/1467-8535.00288

Petrides, L. (2006). Data use and school reform. *T.H.E. Journal*, *33*(8), 38–41.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature review of e-assessment*. Bristol, UK: Nesta Future Lab.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195152968.001.0001

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage Publications.

ADDITIONAL READING

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box*. London, England: Nelson Publishing Company.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–149.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Research, theory and practice*. New York, NY: Teachers College Press.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481. doi:10.3102/00346543058004438

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*, 199–208. PMID:3792417

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. doi:10.1037/0033-2909.119.2.254

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice*, *28*(3), 32–37. doi:10.1111/j.1745-3992.2009.00152.x

KEY TERMS AND DEFINITIONS

Covariation: Variation or variance that is correlated between two or more variables.

E-Learning: Learning that uses electronic technology or media (e.g., the Internet) to access education outside of the traditional brick and mortar classroom.

Formative Assessment: Formal and informal assessment methods conducted by educators concurrent with student learning used to adapt teaching and learning activities to improve student achievement.

Multilevel Modeling: Statistical models (e.g., generalizations of linear models such as linear regression) of parameters that vary at more than one level (e.g., nested data).

Proficiency Test: An exam that evidences how competent or skilled a student or learner is in a particular activity or field of study.

Summative Assessment: Assessment methods that are used to evaluate student learning and/or achievement at the end of an instructional cycle.

Time-Varying Covariate: This statistical term (also called a time-dependent covariate) is used in multilevel growth modeling or survival analysis, and indicates that a covariate in the growth model is not constant throughout.