Increasing the Consequential Validity of Reading Assessment Using Dynamic Measurement Modeling: A Comment on Dumas and McNeish (2017)

Denis G. Dumas[1] and Daniel M. McNeish[2]

Dynamic measurement modeling (DMM) has been shown to improve the consequential validity of longitudinal mathematics assessment in the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) database. Here, the authors demonstrate the capability of DMM to similarly improve the consequential validity of ECLS-K reading assessment through the estimation of student-specific learning capacities and growth parameters for reading.

Keywords: achievement gap; longitudinal studies; psychometrics; reading; social class

In the current U.S. educational system, student scores from single-administration assessments are often interpreted not only as pertaining to past and present student performance but also as indicators of student potential to learn in the domain being assessed. Such a conflation of students' current ability to respond to an assessment and their future learning capacity represents a major threat to the consequential validity of any testing program, especially when existing socioeconomic differences among students mean that they begin school with differing levels of academic preparation. In contrast to this nearly ubiquitous single-timepoint static testing paradigm, *dynamic assessment* (DA) incorporates multiple testing occasions separated by instruction (Tzuriel, 2001). The longitudinal scores collected across a DA session enable researchers to make much richer inferences about student learning capacity than would be possible with static scores and therefore greatly increase the consequential validity of measurement (Sternberg et al., 2002). However, the resource-intensive nature of DA has meant that it cannot be widely adopted within U.S. schools.

To apply theoretical principles of DA to large-scale longitudinal testing programs in the United States, McNeish and Dumas (2017) posited a nonlinear growth modeling framework termed *dynamic measurement modeling* (DMM) that is capable of estimating quantities associated with DA (e.g., learning capacity) from existing longitudinal databases. After this initial statistical formulation, Dumas and McNeish showed that DMM was capable of increasing the consequential validity of mathematics scores within the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) data set. Specifically, this increase in consequential validity arose from the reduction of the effect of socioeconomic status (SES), race, and gender between ECLS-K static mathematics scores and DMM estimated mathematics capacities. In this way, DMM was methodologically capable of demonstrating that although students from disadvantaged socioeconomic backgrounds score lower than their more privileged peers on static math tests, they retain equal capacity to learn math in the future.

However, it is not yet known whether the efficacy of DMM to improve consequential validity of assessment is limited to the domain of mathematics or whether it may extend to other commonly measured constructs such as reading ability. Specifically, in previous work (Dumas & McNeish, 2017), the J-shaped nonlinear growth trajectory of ECLS-K mathematics scores allowed for the estimation of student-specific upper asymptotes on mathematics learning, and those asymptotic estimates were used as learning capacity scores. This nonlinear functional form of student learning trajectories in mathematics was integral to the success of DMM to improve the consequential validity of mathematics assessment because individual differences in the length of time (i.e., number of years since kindergarten) that students were capable of learning at their most rapid were not associated with student SES, race, or gender. However, such a phenomenon may or may not be observed with ECLS-K reading scores, especially if initial reading ability at kindergarten is strongly positively

[1]University of Denver, Denver, CO
[2]Arizona State University, Tempe, AZ

Educational Researcher, Vol. XX No. X, pp. 1–3
DOI: 10.3102/0013189X18797621
Article reuse guidelines: sagepub.com/journals-permissions

related to the rate at which students can develop further reading skills (i.e., a Matthew effect; Pfost, Hattie, Dörfler, & Artelt, 2014). In contrast, if the development of reading ability follows a similarly J-shaped nonlinear trajectory to mathematics ability, then asymptotic capacity scores for reading may be estimated for every student, and those scores could potentially be unassociated with demographic variables in the same way as in prior DMM work with mathematics data. In this brief report, we specifically investigate the capability of DMM to estimate student capacities from ECLS-K reading scores and explicitly test for consequential validity differences between static reading ability scores and DMM estimated reading capacities following the same methodological steps as Dumas and McNeish (2017).

Fitting a DMM to ECLS-K Reading Scores

Seven timepoints of test score data from the ECLS-K database were utilized in this analysis: fall and spring of kindergarten, fall and spring of Grade 1, spring of Grade 3, spring of Grade 5, and spring of Grade 8. These reading scale scores were vertically scaled across timepoints. As with ECLS-K mathematics, the DMM fit to these reading scores featured a J-shaped growth trajectory, meaning that the growth in student reading ability occurred most rapidly in the beginning of students' academic development, and student growth-rate slowed as schooling progressed, eventually approaching an upper asymptote. This nonlinear growth trajectory of ECLS-K reading scores for the *i*th student in the set at the *t*th timepoint can be formalized as:

$$Reading_{it} = \beta_{0i} + \frac{(\beta_{Ui} - \beta_{0i})Time}{\beta_{Ri} + Time} + d_{it}, \tag{1}$$

where β_{0i} represents the students' initial value when time is 0 (i.e., kindergarten fall), β_{Ui} represents the upper asymptote on student reading ability growth (i.e., learning capacity for ECLS-K reading), β_{Ri} represents rate parameter that estimates the point in time when ability growth is halfway between the initial value and asymptote, and d_{it} is a residual term. Additionally, each of the beta coefficients in Equation 1 are composed of a population-averaged fixed effect (α) and a student-specific random effect (ζ_i),

$$\begin{aligned} \beta_{0i} &= \alpha_0 + \zeta_{0i} \\ \beta_{Ui} &= \alpha_U + \zeta_{Ui}. \\ \beta_{Ri} &= \alpha_R + \zeta_{Ri} \end{aligned} \tag{2}$$

These random effects (ζ_i) allow each individual student in the data set to have their own unique reading growth trajectory. To visualize these student-specific J-shaped growth trajectories and asymptotes, Figure 1 includes two DMM reading trajectory plots of 25 randomly drawn students from the ECLS-K data set, with the sample mean trajectory plotted in bold as a comparison.

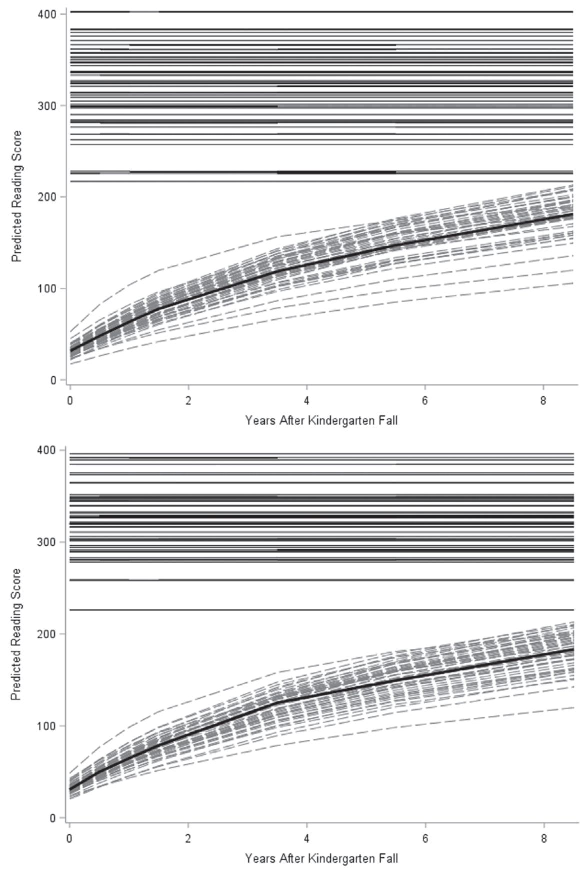

FIGURE 1. *Dynamic measurement modeling reading trajectory plots for two random samples of 25 students from the Early Childhood Longitudinal Study-Kindergarten data set, with superimposed sample mean trajectory (bold) and capacity estimates (horizontal lines).*

Reliability of DMM Estimates

Conditional reliability formulas for item response theory models from Nicewander (2018) can be adapted for use in the DMM context by substituting the appropriate information function. This allows reliability to vary across the range of possible capacity estimates; the conditional reliability formula for DMM reduces to:

$$\rho_{XX'}|\beta_{Ui} = 1 - \frac{Var(\hat{\beta}_{Ui})}{Var(\zeta_{Ui})}, \tag{3}$$

where the reliability at the value of β_{Ui} is equal to one minus the ratio of the sampling variability at β_{Ui} and the variance of all capacity random effects. This procedure returns a reliability estimate for possible values of β_{Ui}, and integrating across all values of β_{Ui} returns a marginal reliability that summarizes the conditional reliability across the distribution of capacity scores. For ECLS-K reading scores, this marginal reliability estimate was .54. Following previous work in which the measurement of reading tends to display somewhat lower reliability than the measurement of mathematics (e.g., Carpenter & Paris, 2005; Clemens et al., 2015; Weiland et al., 2012), this coefficient is lower than the marginal reliability of ECLS-K mathematics capacities, which was .68.

Consequential Validity of DMM Estimates

After saving DMM capacity estimates for each student in the ECLS-K data set, general linear models (GLMs) were used to predict these capacity scores from student race/ethnicity, gender, and

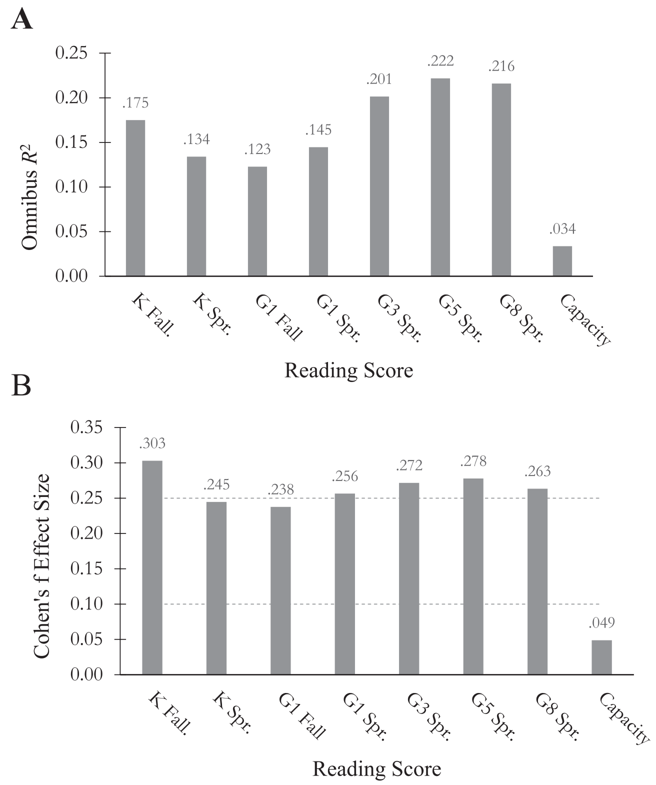

FIGURE 2. *Effect sizes plots.*
Note. (A) Omnibus R2 values showing the amount of total variation explained in reading scale scores and dynamic measurement modeling (DMM) capacity estimates by gender, race/ethnicity, socioeconomic status (SES), and all two and three-way interactions. (B) The effect size (Cohen's f) of SES on Early Childhood Longitudinal Study-Kindergarten reading scale scores and DMM capacity estimates. The dashed horizontal line at .10 represents the cutoff for a "small" effect; the dashed line at .25 represents the cutoff for a "medium" effect.

SES (which was formulated using a principal-component in the same way as in Dumas & McNeish, 2017). The two- and three-way interactions among these predictors were also entered into the GLMs. Effect size estimates from these GLMs are depicted in Figure 2. As can be seen, the omnibus R^2 from each of these models are generally much stronger when predicting single-timepoint ability scores than when predicting DMM capacity scores. Additionally, the specific effect size (i.e., Cohen's f) associated with the SES predictor hovers around "medium" strength (i.e., $f = .25$) across each of the single-timepoint ability scores but drops by a sizeable margin when predicting DMM capacity scores ($f = .05$). Specifically, the effect of SES on DMM estimated reading capacity scores is less than half of the accepted cutoff for a "small" effect, indicating its negligibility. In this way, DMM appears capable of improving the consequential validity of ECLS-K reading scores in the same way it did with ECLS-K mathematics scores.

REFERENCES

Carpenter, R. D., & Paris, S. G. (2005). Issues of validity and reliability in early reading assessments. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 279–304). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Clemens, N. H., Hagan-Burke, S., Luo, W., Cerda, C., Blakely, A., Frosch, J., . . . Jones, M. (2015). The predictive validity of a computer-adaptive assessment of kindergarten and first-grade reading skills. *School Psychology Review*, *44*(1), 76–97. doi:10.17105/SPR44-1.76–97

Dumas, D., & McNeish, D. (2017). Dynamic measurement modeling: Using non-linear growth models to estimate student learning capacity. *Educational Researcher*, *46*, 284–292. doi:10.3102/0013189X17725747

McNeish, D., & Dumas, D. (2017). Non-linear growth models as measurement models: A second-order growth curve model for measuring potential. *Multivariate Behavioral Research*, *52*, 61–85.

Nicewander, W. A. (2018). Conditional reliability coefficients for test scores. *Psychological Methods*, *23*(2), 351–362.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research*, *84*(2), 203–244. doi:10.3102/0034654313509492

Tzuriel, D. (2001). *Dynamic assessment of young children*. New York, NY: Kluwer Academic.

Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., . . . Bundy, D. A. (2002). Assessing intellectual potential in rural Tanzanian school children *Intelligence*, *30*, 141–162.

Weiland, C., Wolfe, C. B., Hurwitz, M. D., Clements, D. H., Sarama, J. H., & Yoshikawa, H. (2012). Early mathematics assessment: Validation of the short form of a prekindergarten and kindergarten mathematics measure. *Educational Psychology*, *32*(3), 311–333. doi:10.1080/01443410.2011.654190

AUTHORS

DENIS G. DUMAS, PhD, is assistant professor of research methods and statistics at The University of Denver Morgridge College of Education, 1999 East Evans Avenue, Denver CO, 80208; *denis.dumas@du.edu*. In his research, he works to understand student learning trajectories through the development and application of nonlinear growth models.

DANIEL M. McNEISH, PhD, is assistant professor of quantitative psychology at Arizona State University, PO Box 871104, Department of Psychology, Tempe, AZ, 85287; *dmcneish@asu.edu*. He focuses on statistical methods for correlated data and challenging data structures such as small samples and missing data.

Manuscript received April 17, 2018
Revision received July 15, 2018
Accepted July 23, 2018