Validating a Measure of Educators' Knowledge of Reading

Previous research has found that teachers' reading knowledge, as measured directly by a

teacher knowledge assessment, is significantly linked to students' achievement in reading

comprehension (Kelcey, 2011). However, the limited scope of instruments for assessing

teachers' knowledge of reading has been noted for at least 40 years (Narang, 1978), despite

ongoing concerns about the capacity of educators to provide the kinds of comprehensive reading

instruction deemed necessary for helping most students become successful readers (McCombes-

Tolis & Feinn, 2008). Evaluation of these instruments is important to ensure trust in their

reliability and validity as well as gauge the appropriateness of their use in measuring teachers'

literacy knowledge. In this study, we explored the psychometric characteristics of the Teacher

Knowledge of Early Literacy Skills (TKELS) test developed by the Regional Educational

Laboratory (REL) Southeast to measure teachers' knowledge of reading (Folsom et al., 2017).

**TKELS Analysis**

The TKELS consists of 4-option, multiple-choice items targeting content knowledge of

early literacy skills, including scenarios related to teaching those skills. Each item has a single

correct response. There were two forms of the test (Forms A and B), each consisting of 31 items

with 8 common items between the forms. Raw scores for participants who completed all items

on one of the forms were converted into a TKELS scale score with a range of 20 to 80.

Pre- and in-service educators from various K-12 school districts, enrichment programs,

and teacher preparation programs at institutes of higher education ($N = 469$) were randomly

assigned by the computer delivery system to take one of the two forms (Form A = 235; Form B =

234) during each administration period. In some cases, participants took a pre- and posttest as

part of a professional learning experience, but other participants took the TKELS only once. The

data from all pretest administrations and participants with only one administration were considered "time 1." Data from all posttest administrations were considered "time 2."

**Results**

As shown in Table 1, similar scale score and Coefficient Alpha values were found for both forms, and histograms revealed approximately normal distributions of scores across forms.

**Table 1**

*Summary of CTT for Form A and Form B*

| TKELS Form | $N$ | Average TKELS Scale Score | Coefficient Alpha of all 31 items | Coefficient Alpha after removing poor items |
|---|---|---|---|---|
| Form A | 235 | 53.51 | 0.7218 | 0.733 |
| Form B | 234 | 54.87 | 0.7195 | 0.724 |

Prior to fitting IRT models to the item response data, we first assessed and confirmed that the assumptions were not being violated (measurement invariance, unidimensionality). However, ICCs indicated that the test forms had a mixture of good and problematic items, so we also fit individual item information curves (IIC) and test information curves (TIC) to each form. For both forms, the information peaked sharply near ability $\theta = -2$ and was mostly concentrated around that ability. This indicated that the degree of precision at other abilities, especially higher abilities, was fairly low compared to the precision around $\theta = -2$. We subsequently removed the items having very low item discriminations or extreme item difficulties, leaving 21 items remaining on Form A and 20 items on Form B. Improved Coefficient Alpha values are shown in the final column of Table 1.

We then reevaluated test information for each form with these items removed, but information was still highly concentrated at the lower end of the ability scale. This concentrated ability might suit well for a test needed to make a decision about an examinee's proficiency using an ability cut-score. However, for a test such as TKELS that attempts to ascertain teachers' reading knowledge along a continuum, it would be preferred to have a more balanced TIF covering a wider range of abilities (de Ayala, 2009; McDonald, 1999). Thus, a new instrument may be necessary.

In the last stage of our analyses, we fit two-dimensional compensatory multidimensional item response theory (MIRT) models for both Forms A and B, allowing items to load onto both of the latent dimensions. The results of these models and contour plots showed a similar mixture of good- and poor-performing items that were found within the univariate framework.

**Summary**

Our investigation raised important questions regarding the psychometric characteristics of the TKELS. Our recommendation is that the mesaure not be used for making inferences about teachers' knowledge. There is a need to develop psychometrically sound teacher knowledge measures that could inform the professional learning of current and future educators. Given that the TKELS development was funded by the Institute of Education Sciences, there is a risk to co-PIs Reed and Aloe in disseminating the results of the present study. In additon, we are concerned that taking away tests of teachers' knowledge, such as TKELS, without yet having a stronger replacement for the measures will create anxiety among educators. Therefore, we do not plan to publish or present findings at this time but will pursue developing a new test with stronger technical adequacy.

References

de Ayala, R.J. (2009). *The theory and practice of item response theory.* New York, NY: The

Guilford Press.

Folsom, J. S., Smith, K. G., Burk, K., & Oakley, N. (2017). *Educator outcomes associated with*

*implementation of Mississippi's K-3 early literacy professional development initiative.*

REL 2017-270. Washington, DC: U.S. Department of Education, Institute of Education

Sciences, National Center for Education Evaluation and Regional Assistance, Regional

Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs

Kelcey, B. (2011). Assessing the effects of teachers' reading knowledge on students' achievement

using multilevel propensity score stratification. *Educational Evaluation and Policy*

*Analysis*, *33*, 458–482.  doi:10.3102/0162373711415262

McCombes-Tolis, J., & Feinn, R. (2008). Comparing teachers' literacy-related knowledge to

their state's standards for reading. *Reading Psychology, 29*, 236-265. doi:

10.1080/02702710801982258

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Lawrence Erlbaum

Associates.

Narang, H. I. (1978). Measurement of teacher knowledge of reading. *Reading Horizons, 18*, 116-

123.