A Simulated Retest Method for Estimating Classification Reliability

W. Jake Thompson & Amy K. Clark

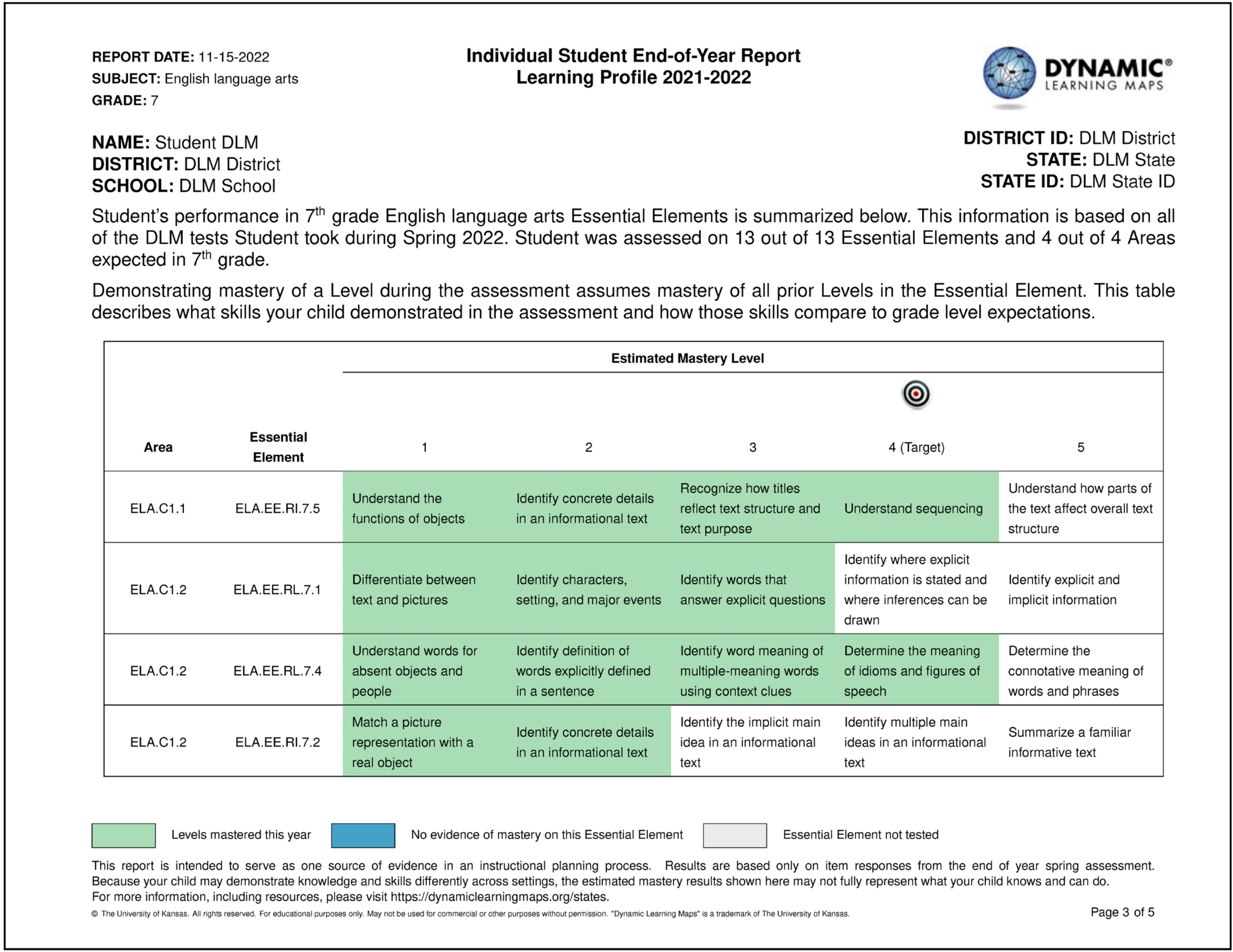
Accessible Teaching, Learning, and Assessment Systems

University of Kansas

DYNAMIC®
LEARNING MAPS

Motivating Example

- Example score report for a DCM-based assessment
- Mastery or proficiency of distinct skills
- Actionable feedback for stakeholders



REPORT DATE: 11-15-2022
SUBJECT: English language arts
GRADE: 7

Individual Student End-of-Year Report
Learning Profile 2021-2022

DYNAMIC® LEARNING MAPS

NAME: Student DLM
DISTRICT: DLM District
SCHOOL: DLM School

DISTRICT ID: DLM District
STATE: DLM State
STATE ID: DLM State ID

Student's performance in 7th grade English language arts Essential Elements is summarized below. This information is based on all of the DLM tests Student took during Spring 2022. Student was assessed on 13 out of 13 Essential Elements and 4 out of 4 Areas expected in 7th grade.

Demonstrating mastery of a Level during the assessment assumes mastery of all prior Levels in the Essential Element. This table describes what skills your child demonstrated in the assessment and how those skills compare to grade level expectations.

		Estimated Mastery Level				
Area	**Essential Element**	1	2	3	4 (Target)	5
ELA.C1.1	ELA.EE.RI.7.5	Understand the functions of objects	Identify concrete details in an informational text	Recognize how titles reflect text structure and text purpose	Understand sequencing	Understand how parts of the text affect overall text structure
ELA.C1.2	ELA.EE.RL.7.1	Differentiate between text and pictures	Identify characters, setting, and major events	Identify words that answer explicit questions	Identify where explicit information is stated and where inferences can be drawn	Identify explicit and implicit information
ELA.C1.2	ELA.EE.RL.7.4	Understand words for absent objects and people	Identify definition of words explicitly defined in a sentence	Identify word meaning of multiple-meaning words using context clues	Determine the meaning of idioms and figures of speech	Determine the connotative meaning of words and phrases
ELA.C1.2	ELA.EE.RI.7.2	Match a picture representation with a real object	Identify concrete details in an informational text	Identify the implicit main idea in an informational text	Identify multiple main ideas in an informational text	Summarize a familiar informative text

🟩 Levels mastered this year 🟦 No evidence of mastery on this Essential Element ⬜ Essential Element not tested

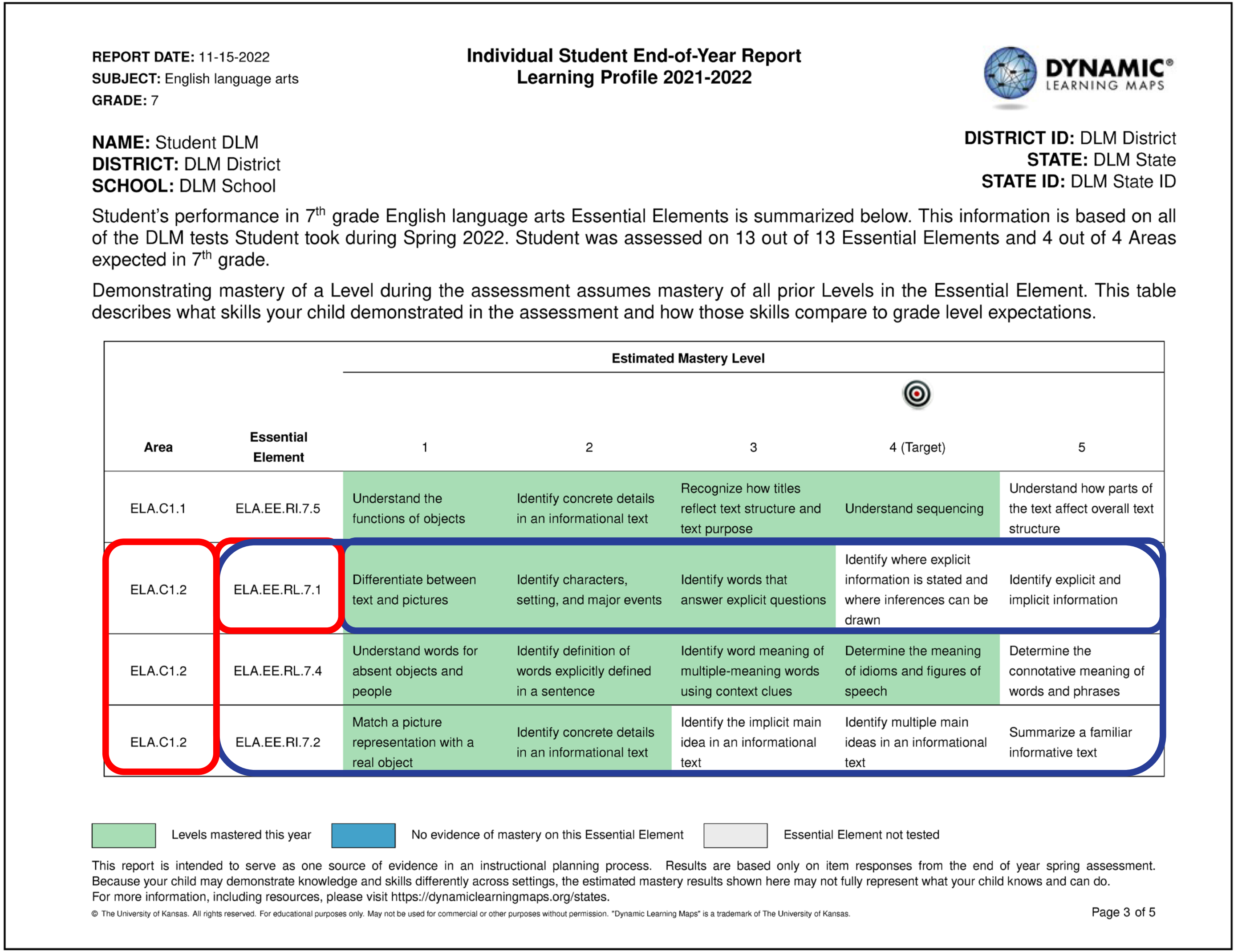
This report is intended to serve as one source of evidence in an instructional planning process. Results are based only on item responses from the end of year spring assessment. Because your child may demonstrate knowledge and skills differently across settings, the estimated mastery results shown here may not fully represent what your child knows and can do. For more information, including resources, please visit https://dynamiclearningmaps.org/states.

Reliability for Diagnostic Assessments

- Well developed methods for evaluating classification accuracy and consistency for diagnostic assessments
  - See Sinharay & Johnson's (2019) *Measures of agreement: Reliability, classification accuracy and classification consistency*
- Focus classification level (i.e., the attribute)
- Operational programs may have other reporting needs

Nested Attributes

- Distinct skills nested within standards
- Further nesting by strand or subjects



REPORT DATE: 11-15-2022
SUBJECT: English language arts
GRADE: 7

Individual Student End-of-Year Report
Learning Profile 2021-2022

DYNAMIC® LEARNING MAPS

NAME: Student DLM
DISTRICT: DLM District
SCHOOL: DLM School

DISTRICT ID: DLM District
STATE: DLM State
STATE ID: DLM State ID

Student's performance in 7th grade English language arts Essential Elements is summarized below. This information is based on all of the DLM tests Student took during Spring 2022. Student was assessed on 13 out of 13 Essential Elements and 4 out of 4 Areas expected in 7th grade.

Demonstrating mastery of a Level during the assessment assumes mastery of all prior Levels in the Essential Element. This table describes what skills your child demonstrated in the assessment and how those skills compare to grade level expectations.

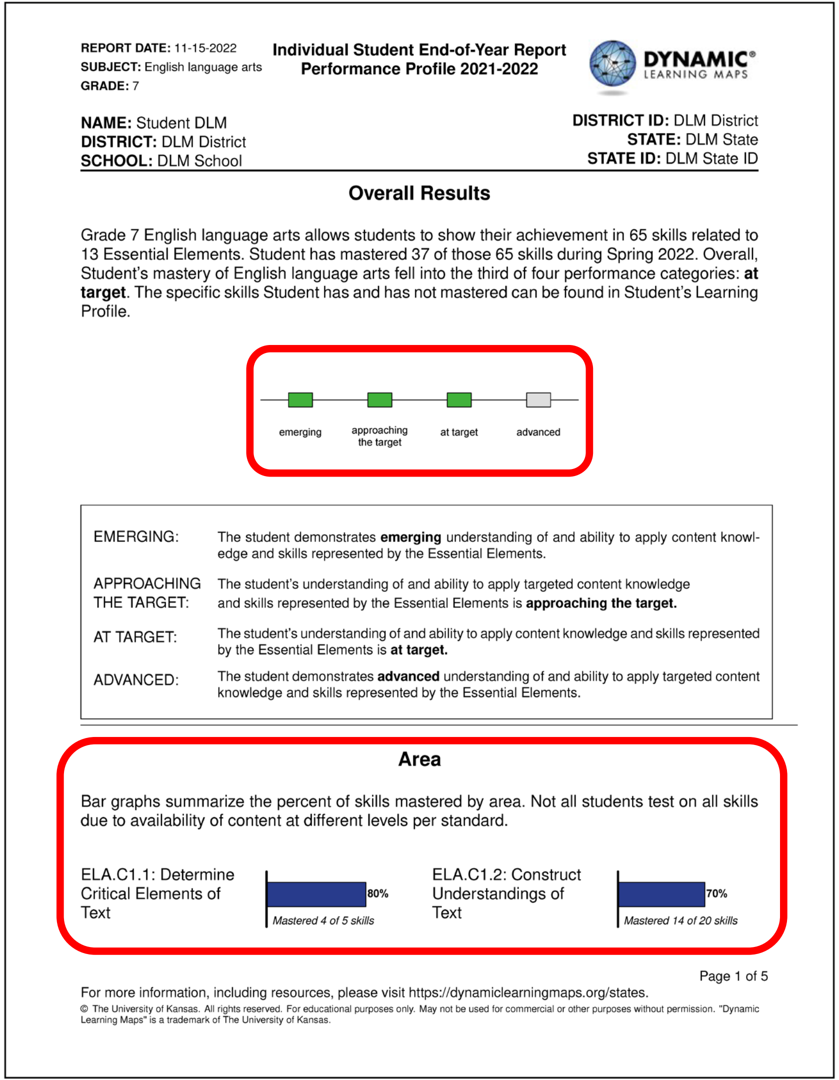
		Estimated Mastery Level				
Area	Essential Element	1	2	3	◎ 4 (Target)	5
ELA.C1.1	ELA.EE.RI.7.5	Understand the functions of objects	Identify concrete details in an informational text	Recognize how titles reflect text structure and text purpose	Understand sequencing	Understand how parts of the text affect overall text structure
ELA.C1.2	ELA.EE.RL.7.1	Differentiate between text and pictures	Identify characters, setting, and major events	Identify words that answer explicit questions	Identify where explicit information is stated and where inferences can be drawn	Identify explicit and implicit information
ELA.C1.2	ELA.EE.RL.7.4	Understand words for absent objects and people	Identify definition of words explicitly defined in a sentence	Identify word meaning of multiple-meaning words using context clues	Determine the meaning of idioms and figures of speech	Determine the connotative meaning of words and phrases
ELA.C1.2	ELA.EE.RI.7.2	Match a picture representation with a real object	Identify concrete details in an informational text	Identify the implicit main idea in an informational text	Identify multiple main ideas in an informational text	Summarize a familiar informative text

🟩 Levels mastered this year 🟦 No evidence of mastery on this Essential Element ⬜ Essential Element not tested

This report is intended to serve as one source of evidence in an instructional planning process. Results are based only on item responses from the end of year spring assessment. Because your child may demonstrate knowledge and skills differently across settings, the estimated mastery results shown here may not fully represent what your child knows and can do. For more information, including resources, please visit https://dynamiclearningmaps.org/states.

Multiple Levels of Aggregation

- Results may be reported as aggregations of classifications
  - E.g., strands or overall performance level



REPORT DATE: 11-15-2022
SUBJECT: English language arts
GRADE: 7

Individual Student End-of-Year Report
Performance Profile 2021-2022

DYNAMIC® LEARNING MAPS

NAME: Student DLM
DISTRICT: DLM District
SCHOOL: DLM School

DISTRICT ID: DLM District
STATE: DLM State
STATE ID: DLM State ID

Overall Results

Grade 7 English language arts allows students to show their achievement in 65 skills related to 13 Essential Elements. Student has mastered 37 of those 65 skills during Spring 2022. Overall, Student's mastery of English language arts fell into the third of four performance categories: **at target**. The specific skills Student has and has not mastered can be found in Student's Learning Profile.

🟩	🟩	🟩	⬜
emerging	approaching the target	at target	advanced

EMERGING:	The student demonstrates **emerging** understanding of and ability to apply content knowledge and skills represented by the Essential Elements.
APPROACHING THE TARGET:	The student's understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements is **approaching the target.**
AT TARGET:	The student's understanding of and ability to apply content knowledge and skills represented by the Essential Elements is **at target.**
ADVANCED:	The student demonstrates **advanced** understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements.

Area

Bar graphs summarize the percent of skills mastered by area. Not all students test on all skills due to availability of content at different levels per standard.

ELA.C1.1: Determine Critical Elements of Text	**80%**	*Mastered 4 of 5 skills*
ELA.C1.2: Construct Understandings of Text	**70%**	*Mastered 14 of 20 skills*

Page 1 of 5

For more information, including resources, please visit https://dynamiclearningmaps.org/states.

Limitations of Current Practice

- *Standards for Educational and Psychological Measurement*
  - 2.3: For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.
- Existing methods do not allow for the aggregation of reliability estimates of distinct skills into an aggregated reliability metric

SIMULATED RETESTS

Overview

- Using estimated model parameters, simulate new responses to assessment items
- Score the simulated assessment using operational scoring rules (e.g., aggregation)
- Compare results from the simulated retest to the observed data
- Reliability is the degree of agreement between observed and simulated results

Step 1: Sample a Student Record

Student	Item 1	Item 2	Item 3	Item 4	Item 5	…
Jayden	1	1	0	1	1	…
Dibanhi	1	1	1	0	0	…
Macyn	1	0	1	1	0	…
Aaron	1	1	1	1	0	…
Kiara	0	1	1	0	1	…
Paulo	0	1	0	1	0	…
Leila	1	1	1	0	0	…
David	0	0	1	1	0	…

Step 2: Simulate a Retest

- Using Paulo's estimated classification probabilities and the model parameters, simulate new item responses
  - E.g., Roussos et al. (2007)
  - Parallel administration using the same items, or
  - Simulation can account for new items (e.g., routing decisions, item selection)

Item	Observed	Simulated
Item 1	0	0
Item 2	1	1
Item 3	0	1
Item 4	1	1
Item 5	0	0
…	…	…

Step 3: Score Simulated Retest

- Using operational scoring rules, score the simulated retest
  - E.g., overall performance level
- Any result calculated from observed data can be calculated from simulated retests (e.g., Clark et al., 2017; Skaggs et al., 2016)

Student	Observed	Simulated
Paulo_1	3	4
…	…	…

Step 4: Repeat

- Draw another student and repeat the process
  - Drawn with replacement
  - Similar to bootstrap sampling (Efron, 2000)
- Sampling will depend on the structure of the assessment
  - Sample 1,000,000 students
  - Sample each student 100 times

Student	Observed	Simulated
Paulo_1	3	4
Aaron_1	3	3
Kiara_1	1	1
Macyn_1	2	2
Aaron_2	3	3
Paulo_2	3	3
Jayden_1	4	3
…	…	…

DYNAMIC® LEARNING MAPS

Step 5: Estimate Reliability

- Calculate appropriate measures of agreement between observed and simulated scores
  - Binary classifications: percent agreement, tetrachoric correlation, Cohen's kappa
  - Polytomous classifications: percent agreement, polychoric correlation, Cohen's kappa
  - Interval scales: Pearson correlation
- May choose to report multiple metrics

Simulated Retest Method is Accurate

- Retest estimates of attribute-level classification accuracy and consistency are nearly identical to non-simulation approaches
- Limited to comparisons at the attribute level (no aggregated comparison metric)

Thompson et al. (2023): *Using simulated retests to estimate the reliability of diagnostic assessment systems.*

Simulated Retest Method is Flexible

- Simulated retests are not limited to attribute-level summaries of reliability
  - Content standard or content strand
- Flexible enough to accommodate any operational scoring rules

Thompson et al. (2019): *Measuring the reliability of diagnostic classifications at multiple levels of reporting.*

DYNAMIC® LEARNING MAPS

Considerations

- For multiple reporting structures, simulated retests offer a straightforward method for assessing reliability
  - If only reporting attribute-level results, simulated retests may not be optimal (i.e., time and computationally intensive)
- Important to evaluate model fit, as the simulation uses the estimated model parameters
- Different summary statistics may be preferred in different contexts
  - Cohen's kappa may be suboptimal with unbalanced classes

Conclusions

- As diagnostic models move from theory to implementation, existing methods for providing technical evidence may need to be adapted for operational settings
- Reliability is one example where existing methods were limiting for operational use
  - Simulated retests overcome this limitation
- Additional work likely needed in other areas
  - E.g., DIF, equating, growth

Get in Touch!

https://dynamiclearningmaps.org/publications

🌐 atlas.ku.edu

✉ atlas-aai@ku.edu

in company/atlas-ku

/ @atlas4learning

🌐 wjakethompson.com

✉ wjakethompson@ku.edu

in in/wjakethompson

// @wjakethompson

DYNAMIC®
LEARNING MAPS

References

Clark, A. K., Nash, B., Karvonen, M., & Kingston, N. (2017). Condensed mastery profile method for setting standards for diagnostic assessment systems. *Educational Measurement: Issues and Practice*, *36*(4), 5–15. https://doi.org/10.1111/emip.12162

Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association*, *95*(452), 1293–1296. https://doi.org/10.2307/2669773

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. L. (2007). The fusion model skills diagnosis system. In J. Leighton & M. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education: Theory and Applications* (pp. 275–318). Cambridge University Press. https://doi.org/10.1017/CBO9780511611186.010

Sinharay, S., & Johnson, M. S. (2019). Measures of agreement: Reliability, classification accuracy, and classification consistency. In M. von Davier & Y.-S. Lee (Eds.), *Handbook of Diagnostic Classification Models* (pp. 359–377). Springer International Publishing. https://doi.org/10.1007/978-3-030-05584-4_17

Skaggs, G., Hein, S. F., & Wilkins, J. L. M. (2016). Diagnostic profiles: A standard setting method for use with a cognitive diagnostic model. *Journal of Educational Measurement*, *53*(4), 448–458. https://doi.org/10.1111/jedm.12125

Thompson, W. J., Clark, A. K., & Nash, B. (2019). Measuring the reliability of diagnostic mastery classifications at multiple levels of reporting. *Applied Measurement in Education*, *32*(4), 298–309. https://doi.org/10.1080/08957347.2019.1660345

Thompson, W. J., Nash, B., Clark, A. K., & Hoover, J. C. (2023). Using simulated retests to estimate the reliability of diagnostic assessment systems. *Journal of Educational Measurement*. https://doi.org/10.1111/jedm.12359