# Establishing Learning Progression-based Vertical Scales to Measure Growth

Lei Yu
Measured Progress

Wonsuk Kim
Measured Progress

Jennifer Dunn
National Center for the Improvement of Educational Assessment

Paper presented as part of the coordinated session "*Modeling Learning Progressions and Informing Practice*" at the annual meeting of National Council on Measurement Education held between April 26 - 30, San Antonio, TX.

## Abstract

A key shift in the common core state standards (CCSS) that started the next generation of assessments is the use of the learning progression-based principles that postulate a staircase of increasing complexity and progressive development of skills as students move from grade to grade. This vertical trajectory of learning defined by learning progressions captures the essence of learning pathways and supports growth modeling using vertical scales. This paper explored in-depth modeling of a learning progression from perspectives: validating the learning progression and measuring student progress on a vertical scale. For learning progression-based results to be valid and useful, it is imperative to empirically validate learning progressions and examine the congruence between theoretical and research expectations and how students actually learn and progress. Data from a state-wide administration of eMPower Reading assessments$^{TM}$ was used. Learning progressions, built in the assessment series, currently serve the purpose of research only. On-grade operational items representing the learning progression were selected, placed in the matrix fieldtest slots of adjacent grades, and used as vertical linking (VL) items. Concurrent calibration was conducted in WINSTEPS to examine the empirical hierarchy of item positions on the learning progression. To obtain ability estimates as student progress measures, separate calibrations were conducted by anchoring parameters and step values of stable VL items from the concurrent calibration and allowing the unstable items to estimate freely. The study demonstrates a method for validating learning progressions and provided recommendations for improvement and refinement based on empirical findings.

## Introduction

A key shift in the common core state standards (CCSS) that started the next generation of assessments is the use of learning progression-based principles that postulate a staircase of increasing complexity and progressive development of skills as students move from grade to grade (Common Core State Standards, 2010a) and that link topics and thinking across grades to build coherent progressions (Common Core State Standards, 2010b). Learning progressions are "descriptions of the successively more sophisticated ways of thinking about an idea that follow one another as students learn" (Wilson & Bertenthal, 2005, p. 48.). They were also similarly defined as "descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic over a broad span of time" (National Research Council, 2007, p. 219). Learning progressions describe coherent pathways students are likely to take to achieve desired proficiency through stages of progress over an extended period of study. Learning is conceptualized as a sequence of increasing sophistication in thinking, understanding, and competency (of a topic or sub-skill) about a topic, an idea, a skill or a concept within a content domain.

The essence of learning progressions is the vertical development of incremental sophistication that connects knowledge, concepts and skills rather than a series of isolated or disconnected events (Heritage, 2008). This vertical trajectory of learning defined by learning progression captures the essence of learning pathways and supports growth modeling using vertical scales (Briggs & Beck, 2015a; 2015b). In the era of educational accountability under NCLB, vertical scaling was used as a key method to measure student growth from grade to grade. This use of vertical scales as a growth measure has been criticized for distorting growth due to construct shift across grade levels (Reckase & Martineau, 2004; Martineau, 2005). The coherent progressions that connect cautiously learning across grades to build new understanding, knowledge, and skills offer unprecedented opportunities to model growth using vertical scales that overcome the "construct shift" limitation inherent in traditional vertical scales.

In addition to vertical scaling (Briggs & Beck, 2015a; 2015b), methods to model learning progressions include a family of structured constructs models to model constructs and the levels of constructs (Wilson, 2009, 2012). Thissen (2015) extended the work of Briggs and Wilson to connect cognitive diagnostic models to conventional IRT models from a theoretical perspective. The Rasch-based family of models was perceived to be more appropriate for assessments designed on the basis of learning progressions (Briggs & Beck, 2015a; 2015b, Wilson 2012). However, research on modeling learning progressions for operational assessments to inform practice is very limited.

Learning progressions are hypothesis about student's learning pathways based on theory and research on how students learn, think about or master a specific topic and need to be empirically tested for construct validity and ultimately for consequential validity (Corcoran, Mosher, & Rogat, 2009). In discussing the role of learning progressions, Black, Wilson, & Yao (2011) noted that progression in learning should also be reflected in curriculum and used to inform both pedagogy and assessments, including formative and summative assessments. Learning progressions need to be formulated on the basis of professional judgment and sound instructional practices. And most importantly, they should also be validated by data gathered through assessment and the interpretation of student responses. If a learning progression functions differently from expectations, the validity of the results based on the learning progression becomes questionable. Therefore, empirical validation is necessary to improve and refine learning progressions for their results to be useful.

This paper was designed to explore in-depth modeling of learning progressions using eMPower Assessments™, an assessment product aligned to college and career readiness standards and based on current research about what students need to know and be able to do for success beyond high school. Embedded in the design of the assessments are learning progress-based principles. Learning progressions, built in the assessment series, currently serve the purpose of research and are not used for score reporting. The Reading assessments consist of three learning progressions centered on anchor standards. One of them – "Craft and Structure" – covers content strands of Reading Literature and Reading for Information and is the focus of the analyses.

The purpose of the study is two-fold: validating the learning progression and measuring student progress on a vertical scale. The data came from a state-wide administration of the assessments administered to grades 3-8. On-grade operational items representing the learning progression were selected, placed in the matrix fieldtest slots of adjacent grades, and used as vertical linking (VL) items. For grade 3, VL items were placed only on the upper grade (grade 4) and for grade 8, VL items were placed only on the lower grade (grade 7). Minimum sample size was 12,546 for operational items and 1,080 for off-grade items. Concurrent calibration was conducted in WINSTEPS to examine the empirical hierarchy of item positions along the learning progression. To obtain ability estimates as student progress measures, separate calibrations were conducted by anchoring parameters and step values of stable VL items from the concurrent calibration and allowing the unstable items to estimate freely.

Learning progressions can provide information about specific learning goals and inform decisions and interventions that need to be introduced to improve student learning. Assessments based on learning progressions can determine not only where students are at each level of progress but also track their progress over time. For such results to be valid and useful, it is imperative to empirically validate learning progressions and examine the congruence between theoretical and research expectations and how students actually learn and progress. The study demonstrates a method for such validation and offers recommendations to improve and refine learning progressions based on empirical findings.

## Method

### Test Design

The data used in the study came from a state-wide administration of eMPower Assessments™ aligned to college and career readiness standards. Administered to grades 3-8, the Reading tests consisted of items representing six content standards: Reading for Informational Text Craft and Structure, Reading for Informational Text Integration of Knowledge and Ideas, Reading for Informational Text Key Ideas and Details, Reading Literature Craft and Structure, Reading Literature Integration of Knowledge and Ideas, and Reading Literature Key Ideas and Details. For adequate representation of the constructs, learning progressions were built around the anchor standards, which resulted in three learning progress: "Craft and Structure", "Integration of Knowledge and Ideas", and "Key Ideas and Details". Item types on the tests included 1-point multiple-choice (MC) items, 2-point evidence-based selected response (EBSR) items, and constructed-response (CR) items that ranged from 2 to 4 points. The tests consisted of core items administered to each grade and matrix fieldtest (FT) item slots. Numbers of core items and possible score points by learning progression are presented in Table 1. For "Craft and Structure", the minimum number of items was 7 at both grades 4 and 7.

Table 1

*Numbers of Core Items and Possible Score Points by Learning Progression by Grade*

| **Learning Progression** | **# of Core Items** | | | | | | **# of Score Points** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | G3 | G4 | G5 | G6 | G7 | G8 | G3 | G4 | G5 | G6 | G7 | G8 |
| Craft and Structure | 11 | 7 | 10 | 9 | 7 | 9 | 14 | 8 | 11 | 10 | 7 | 16 |
| Integration of Knowledge and Ideas | 7 | 6 | 4 | 8 | 6 | 6 | 10 | 10 | 7 | 10 | 13 | 7 |
| Key Ideas and Details | 8 | 13 | 12 | 9 | 13 | 11 | 11 | 17 | 17 | 17 | 17 | 14 |
| Total | 26 | 26 | 26 | 26 | 26 | 26 | 35 | 35 | 35 | 37 | 37 | 37 |

On-grade core items representing each learning progression were selected and placed in the matrix fieldtest slots of adjacent grades, and used as vertical linking (VL) items. Take "Craft and Structure" for example. Four core items in Grade 4 were vertical linking items. They were administered to both grade 3 and grade 5 students in the FT positions. Numbers of VL items between adjacent grades are presented in Table 2. It can be seen that VL items range from 4 (between grades 4 & 5 and grades 7 & 8 for "Integration of Knowledge and Ideas") to 16 (between grades 7 & 8 for "Key Ideas & Details").

Table 2

*Numbers of Vertical Linking Items between Adjacent Grades by Learning Progression*

| **Learning Progression** | **3/4** | **4/5** | **5/6** | **6/7** | **7/8** |
|---|---|---|---|---|---|
| Craft and Structure | 9 | 10 | 10 | 7 | 10 |
| Integration of Knowledge and Ideas | 8 | 4 | 7 | 7 | 4 |
| Key Ideas and Details | 13 | 13 | 13 | 14 | 16 |

The total number of items administered to each grade by learning progression is included in Table 3. The total number of non-overlapping items for each learning progression is also included.

Table 3

*Total Number of Items Administered to Each Grade by Learning Progression.*

| **Learning Progression** | **G3** | **G4** | **G5** | **G6** | **G7** | **G8** | **Total Non-overlapping Items** |
|---|---|---|---|---|---|---|---|
| Craft and Structure | 21 | 18 | 18 | 20 | 16 | 14 | 59 |
| Integration of Knowledge and Ideas | 14 | 14 | 11 | 12 | 13 | 8 | 42 |
| Key Ideas and Details | 21 | 24 | 27 | 24 | 27 | 19 | 70 |

**Sample**

Data used in the study included students in grades 3-8 who were administered the learning progression based assessments. Sample sizes for the core items ranged from 12,796 to 13,103. As mentioned earlier, vertical linking items were placed in the matrix slots of adjacent grades. Sample sizes for the matrix items ranged from 1093 to 2219.

## Analysis Design

The analysis was designed to explore in-depth modeling of one single learning progression. Concurrent calibration was performed in WINSTEPS to examine the construct measured by the learning progression and the alignment of items on the construct/Wright map. In addition to producing item difficulty and person ability on the same scale, concurrent calibration presented a preferred choice given the limited number of core items and vertical linking items on the learning progress.

The second purpose of the study was to measure student progress on the learning progression on a vertical scale. Conceptually, results from concurrent calibration could be used as students' progress measure. However, best practice was followed in this study by first checking the stability of the VL items using the following procedure:

- Fix person ability estimates from the concurrent calibration
- Conduct separate calibration for each grade using the fixed person ability estimates
- Evaluate the stability of the parameters of the VL items using the criterion of |.3|.

Next, separate calibration was conducted to obtain person ability estimates by fixing on the parameters (and step values) of the stable VL items, allowing the unstable VL items contributing to scoring to be estimated for that grade. Note only on-grade VL items were part of the core items that should contribute to scoring. Core items that were not vertical linking items were also fixed on the parameters from concurrent calibration. Since item parameters obtained from concurrent calibration were on the same scale, the obtained person estimates thereby were also on the same scale. Student performance on the learning progression was then evaluated. For all the analyses, learning progression "Craft and Structure" was used.

# Results

## Validating the Learning Progression

Concurrent calibration was conducted to calibrate all the 59 items on the learning progression of "Craft and Structure", which placed item difficulty and person ability estimates on the same scale. Calibration converged at .001 after 341 cycles, with some apparent struggles. Item statistics from concurrent calibration are presented in Table 4 in Appendix A. The b values have a mean of 0 and the standard deviation (SD) of 1.19, with a range of -2.89 and +3.23. Their standard errors range from .01 to .08. No item has an infit mean square value exceeding 1.30.

**Principal component analysis of standardized residuals**. The Rasch-based principal component analysis of standardized residuals conducted in WINSTEPS is presented in Table 5. The results show the variance explained by the measures, which is the sum of the variance explained by the persons and the variance explained by the items, is 44.30%. The observed variance is essentially the same as the expected variance, which is 43.80%. This clearly indicates unidimensionality of the measured construct. The amount of variance explained by items, which is 24.50%, is higher than the variance explained by persons, which is 19.80%. This is not surprising, since the standard deviation (SD) of items equals 1.18 and is slightly higher than that for persons, which equals 1.03 when non-extreme persons were excluded and 1.14 when they were included.

Table 5
*Standardized Residual Variance*

| Variance | Eigenvalue | Observed | | Expected |
|---|---|---|---|---|
| Total raw variance in observations | 105.91 | 100.00% | | 100.00% |
| Raw variance explained by measures | 46.91 | 44.30% | | 43.80% |
| Raw variance explained by persons | 20.96 | 19.80% | | 19.60% |
| Raw Variance explained by items | 25.95 | 24.50% | | 24.20% |
| Raw unexplained variance (total) | 59.00 | 55.70% | 100.00% | 56.20% |
| Unexplained variance in 1st contrast | 1.35 | 1.30% | 2.30% | |

**The Wright map.** The Wright map or the construct map that aligns item difficulty and person ability estimates is presented in Figure 1. Three observations can be noted. First, items and person abilities were aligned along the measurement scale, indicating the items were targeted and, therefore, appropriate for the intended population. Second, a significant portion of items clustered between 0 and 1 logit on the scale, indicating more separation of items is needed. Finally, the location of some items along the construct is disordered. For example, Item 8M56C, a grade 8 core MC item, has a b value of -1.37 and is, therefore, located at the lower part of the scale while Item 5O34C, a grade 5 core OR item, has a b value of 2.09 and is located at the upper part of the scale. This indicates that some items are not located as expected across grade levels along the defined learning progression.

### Measuring Student Progress on the Learning Progression

It would be ideal to provide measures of student progress on a learning progression with the items targeted at different grades well aligned on the scale. However, for illustrative purposes, analysis was conducted to demonstrate the process for obtaining such measures. Additionally, results from concurrent calibration could be used as students' progress measure. However, best practice was followed in this study by first checking the stability of the VL items.

**Evaluating the stability of linking item parameters.** An important procedure in test equating and linking is evaluating the stability of parameters of linking items, which is also a standard operational practice used for vertical linking. To obtain item parameters by grade on the same scale, person ability estimates from concurrent calibration were fixed and used in separate calibration. With 6 grades, 6 separate calibrations were conducted.

Parameter estimates of vertical linking items by grade are presented in Table 6 in Appendix A. The table also includes differences of item parameters between adjacent grades. Negative values indicate that items are easier for the higher grades, which is expected, and positive values show that items are harder for the higher grades. Using the criterion of b-value difference of |.3|, 7 VL items were flagged with unstable parameters and are shown in italic (and in red). Of these 7 items, 5 became easier, which is in the expected direction, and 2 became more difficult. Three of the 7 items were core items in grade 4.

**Measuring student progress**. Following the evaluation of vertical linking items, the next step was to produce ability estimates for all the students on the learning progression that were on the same vertical scale. To produce ability estimates, core items for each grade which contributed to scoring were fixed on the parameters and step values of items from concurrent calibration. For core items that were also vertical linking items, only parameters of stable items were fixed and the unstable items, on the other hand, were allowed to estimate freely. Take Grade 8 as an example. Item 8M51C, one of the vertical linking items, was identified as showing unstable item parameters in Table 6. In separate calibration, its parameter, therefore, was not fixed on 1.1936 obtained from concurrent calibration, as shown in Table 4; instead, it was freed to estimate. Person ability estimates were computed using this procedure in separate calibration for each grade.

Averages of the obtained theta estimates and their standard errors (SEs) were computed and summarized in Table 7. The results show the average ability estimates increase with the increase of the grade level. The average standard errors range from .68 to 1.03. The two largest SEs are associated with the ability estimates of Grade 4 and Grade 7 where both had 7 items on the learning progression, the lowest of all.

Table 7

*Descriptive Statistics of Ability Estimates and Their Standard Errors*

| Grade | Average Theta | Average SE |
|---|---|---|
| 3 | -0.29 | 0.68 |
| 4 | 0.16 | 1.03 |
| 5 | 0.59 | 0.78 |
| 6 | 0.84 | 0.81 |
| 7 | 1.01 | 0.97 |
| 8 | 1.30 | 0.68 |

Average ability estimates by grade were also plotted in Figure 2. It can be seen that student performance on the learning progression increases with the increase of grade level as expected. The difference between average thetas is .45 for grades 4 and 3, .43 for grades 5 and 4, .25 for grades 6 and 5, .17 for grades 7 and 6, and .29 for grades 8 and 7. It can be seen that students at different grade levels progresses at different rate and more progress is observed in grade 4 and grade 5 than at other grades.

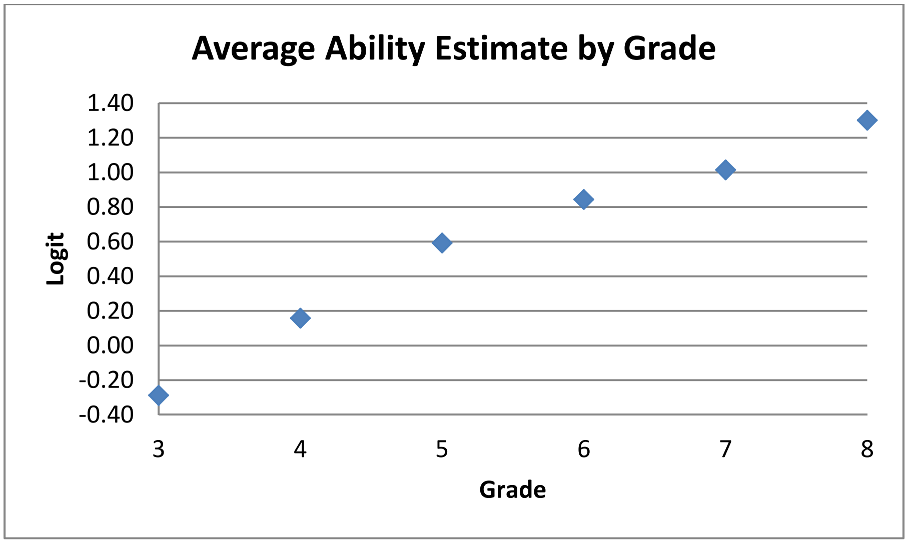


Figure 2. Plot of average ability estimate by grade.

## Summary and Recommendations

### Summary

The study explored in-depth modeling of a learning progression from two perspectives: validating the learning progression and measuring student progress on a vertical scale. Using Craft and Structure as an example, concurrent calibration was conducted to calibrate all the items unto the same scale. Principal component analysis of standardized residual validated the unidimensionality of the constructed measured by the learning progression. The Wright map shows that the items on the learning progression were aligned along the measurement scale and appropriate for the student population. However, some higher grade level items were found substantially easier while some lower grade items were found substantially harder. Their locations on the scale, therefore, were disordered, which was unexpected. This indicates that some lower level items require more sophistication to respond and solve than some higher level items. The empirical hierarchy of item locations on the learning progression differs from what is intended.

To measure student progress on a vertical scale, vertical linking items were first evaluated for their stability. Parameters of vertical linking items by grade were obtained in separate calibration by anchoring person ability estimates from concurrent calibration. Most of the vertical linking items demonstrate stable parameter estimates across grades. However, 7 items do show substantial differences using the criteria of |.3| to flag instability. Of these 7 items, 5 became easier, which is in the expected direction, and 2 became more difficult, indicating reversal. To place student ability estimates on the same scale, separate calibration was conducted for each grade. In separate calibration, item parameters and step values of core items and on-grade stable VL items were fixed on those obtained from the concurrent calibration to produce ability estimates as progress measures. Average ability estimates indicate student performance on the learning progression increases as grade level increases, as is expected. Average standard errors of ability estimates are found to be larger at grade levels where the number of items on the learning progress is small. The competing need for shorter testing time (often met with a shorter test) and the desire for high test reliability and, therefore, small standard error is an eternal challenge at least for the foreseeable future, particularly for learning progression-based assessments.

One thing to note is that concurrent calibration is a preferred choice given the small number of items on the learning progression and the small number of VL items, should non-convergence be an issue. A common practice in evaluating the stability of vertical linking items is removing them from the calibration of the vertical scale. On the other hand, research also shows that item reversal, the fact that items do not get easier from one grade to next, is not a sufficient reason for excluding them from the linking set (Briggs & Dadey, 2015). The procedure used in this study, which fix the parameters of stable items but allow unstable items to be estimated in measured student progress, offers a solution that takes into the performance of such items into account without excluding them.

### Recommendations

The study identified areas of additional investigation and improvement with regard to the learning progression under study. Based on the findings, the following recommendations can be made:

1. Evaluate the content of items whose locations are disordered on the learning progression and different from expectations.
2. Identify other factors (e.g., curriculum and teaching) that may contribute to the observed disordering in conjunction with theory-based expectations.
3. Evaluate item locations within grade levels.
4. Examine the content and statistics of the items clustered in the middle of the scale to identify possibilities to increase the separation of these items and, thereby, increase the amount of explained variance.
5. Administer the tests with the revised and refined learning progressions and conduct additional validation analyses.
6. Refrain from reporting any learning progression-based results until all the learning progressions are validated with the empirical data.

This study provides an in-depth analysis of modeling one learning progression. Future research will be conducted to explore modeling the other two learning progressions that are components of the assessments. In addition, future research can also be conducted to explore modeling multiple learning progressions under the framework of unidimensional item response theory, since it may be desirable to produce and report results on the same scale.

# References

Black, P., Wilson, M., & Yao, S. (2011) Road Maps for Learning: A Guide to the Navigation of Learning Progressions, *Measurement: Interdisciplinary Research and Perspectives*, 9:2-3, 71-123, DOI: 10.1080/15366367.2011.591654

Briggs, D. C. & Dadey, N. (2015). Making sense of common test items that do not get easier over time: Implications for vertical scale designs. *Educational Assessment, 20*(1), 1-22.

Briggs, D. C., & Peck, F. A. (2015a). Using Learning Progressions to Design Vertical Scales that Support Coherent Inferences about Student Growth, *Measurement: Interdisciplinary Research and Perspectives*, 13:2, 75-99, DOI: 10.1080/15366367.2015.1042814.

Briggs, D. C., & Peck, F. A. (2015b). Rejoinder to Commentaries on Using Learning Progressions to Design Vertical Scales, *Measurement: Interdisciplinary Research and Perspectives*, 13:3-4, 206-218, DOI: 10.1080/15366367.2015.1104113.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidenced-basedapproach to reform.* (Consortium for Policy Research in Education Research Report #RR-63). New York, NY: Center on Continuous Instructional Improvement, Teachers College-Columbia University.

Common Core State Standards Initiative. (2016). *Standards in you state*. Retrieved from http://www.corestandards.org/standards-in-your-state/ .

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. Paper prepared for the Formative Assessment for Teachers and Students (FAST) State Collaborative on Assessment and Student Standards (SCASS) of the Council of Chief State School Officers (CCSSO).

National Research Council (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: The National Academy Press.

Martineau, J. A. (2005). *Un-distorting measures of growth: Alternatives to traditional vertical scales*. Paper presented at the 35th Annual Conference of the Council of Chief State School Officers.

Martineau, J. A. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, *31*(1), 35–62.

Reckase, M. D., & Martineau, J. A. (2004, October). *Growth as a Multidimensional Process.* Paper presented at the Annual Meeting of the Society for Multivariate Experimental Psychology, Naples, FL.

Thissen, D., (2015). Growth through levels. Measurement: Interdisciplinary Research and Perspectives, 13(2), 128-131, DOI: 10.1080/15366367.2015.1055146

Wilson, M., & Bertenthal, M. (2005). *Systems for state science assessment*. Washington, DC: National Academy Press.

Wison, M. (2009). The structured constructs model (SCM): A family of statistical models related to learning progressions. Paper presented at the Learning Progressions in Science Conference, June 2009, Iowa City, IA.

Wilson, M. (2012). Responding to a challenge that learning progressions pose to measurement practice. Hypothesized links between dimensions of the outcome progression. In Alonzo, C. A., & Gotwals A. W. (eds.), Learning progressions in Science: Current Challenges and Future Directions, 317-343

## Appendix A

Table 4

*Item Statistics from Concurrent Calibration*

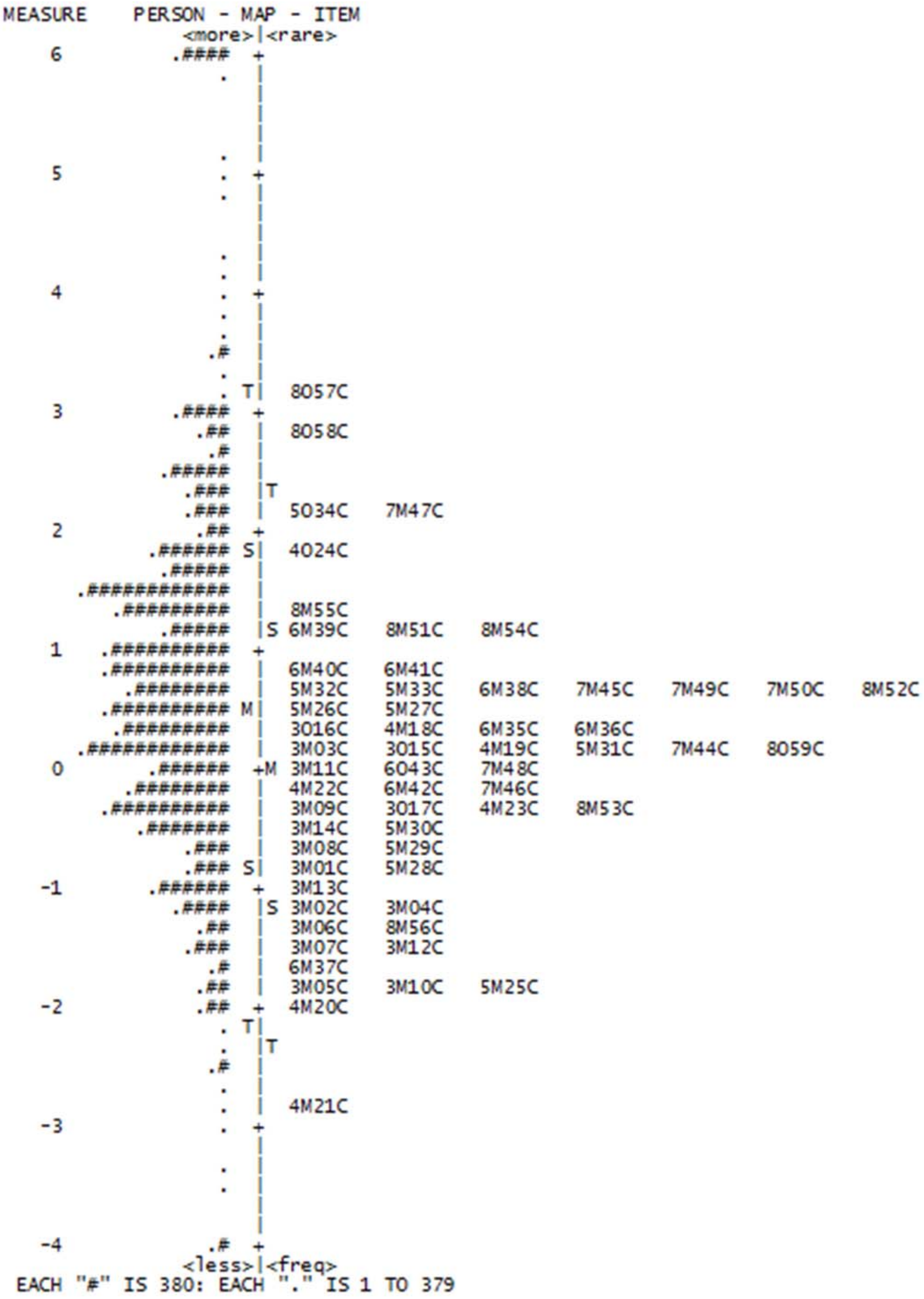
| Item | MEASURE | MODLSE | IN.MSQ | OUT.MS | NAME | Item | MEASURE | MODLSE | IN.MSQ | OUT.MS | NAME |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.8908 | 0.021 | 0.95 | 0.93 | 3M01C | 31 | 0.1749 | 0.0203 | 1.08 | 1.13 | 5M31C |
| 2 | -1.1301 | 0.02 | 0.89 | 0.82 | 3M02C | 32 | 0.7229 | 0.0201 | 1.04 | 1.09 | 5M32C |
| 3 | 0.165 | 0.021 | 1.05 | 1.1 | 3M03C | 33 | 0.65 | 0.0201 | 1.02 | 1.05 | 5M33C |
| 4 | -1.172 | 0.02 | 0.97 | 0.94 | 3M04C | 34 | 2.0914 | 0.0151 | 0.86 | 0.8 | 5O34C |
| 5 | -1.8353 | 0.023 | 0.98 | 0.94 | 3M05C | 35 | 0.2555 | 0.0203 | 1.12 | 1.17 | 6M35C |
| 6 | -1.3245 | 0.022 | 0.89 | 0.84 | 3M06C | 36 | 0.3332 | 0.0187 | 0.99 | 0.98 | 6M36C |
| 7 | -1.4263 | 0.022 | 1 | 1.04 | 3M07C | 37 | -1.6324 | 0.0294 | 0.89 | 0.69 | 6M37C |
| 8 | -0.6306 | 0.02 | 1.02 | 1.03 | 3M08C | 38 | 0.6783 | 0.0185 | 0.97 | 0.96 | 6M38C |
| 9 | -0.2925 | 0.068 | 1.13 | 1.2 | 3M09C | 39 | 1.1027 | 0.0186 | 1.08 | 1.12 | 6M39C |
| 10 | -1.8562 | 0.08 | 0.92 | 0.91 | 3M10C | 40 | 0.8213 | 0.0199 | 1.09 | 1.12 | 6M40C |
| 11 | -0.0351 | 0.069 | 0.95 | 0.94 | 3M11C | 41 | 0.8456 | 0.02 | 1.04 | 1.07 | 6M41C |
| 12 | -1.5155 | 0.077 | 0.9 | 0.78 | 3M12C | 42 | -0.1787 | 0.0213 | 0.91 | 0.86 | 6M42C |
| 13 | -1.0081 | 0.072 | 0.96 | 0.95 | 3M13C | 43 | 0.0689 | 0.0122 | 0.9 | 0.84 | 6O43C |
| 14 | -0.4298 | 0.069 | 1.04 | 1.1 | 3M14C | 44 | 0.1528 | 0.0189 | 0.89 | 0.82 | 7M44C |
| 15 | 0.1805 | 0.014 | 1.1 | 1.12 | 3O15C | 45 | 0.6263 | 0.0182 | 1.09 | 1.14 | 7M45C |
| 16 | 0.3159 | 0.013 | 1.1 | 1.4 | 3O16C | 46 | -0.1794 | 0.0222 | 0.97 | 0.94 | 7M46C |
| 17 | -0.3094 | 0.014 | 0.93 | 0.92 | 3O17C | 47 | 2.1674 | 0.0199 | 1.09 | 1.21 | 7M47C |
| 18 | 0.3666 | 0.019 | 1.07 | 1.14 | 4M18C | 48 | -0.056 | 0.0195 | 1 | 1.02 | 7M48C |
| 19 | 0.1925 | 0.019 | 1.08 | 1.14 | 4M19C | 49 | 0.6199 | 0.0183 | 0.92 | 0.89 | 7M49C |
| 20 | -1.9802 | 0.025 | 0.94 | 0.79 | 4M20C | 50 | 0.7014 | 0.0205 | 1.09 | 1.11 | 7M50C |
| 21 | -2.8898 | 0.034 | 0.89 | 0.63 | 4M21C | 51 | 1.1936 | 0.0193 | 1.27 | 1.53 | 8M51C |
| 22 | -0.181 | 0.021 | 0.99 | 0.98 | 4M22C | 52 | 0.6366 | 0.0207 | 1.04 | 1.07 | 8M52C |
| 23 | -0.2591 | 0.021 | 1 | 1 | 4M23C | 53 | -0.3678 | 0.023 | 0.9 | 0.79 | 8M53C |
| 24 | 1.8733 | 0.018 | 1 | 1.01 | 4O24C | 54 | 1.0877 | 0.0194 | 1.05 | 1.11 | 8M54C |
| 25 | -1.7868 | 0.026 | 0.92 | 0.77 | 5M25C | 55 | 1.4138 | 0.0193 | 0.97 | 1 | 8M55C |
| 26 | 0.5061 | 0.019 | 0.95 | 0.94 | 5M26C | 56 | -1.3723 | 0.0314 | 0.86 | 0.66 | 8M56C |
| 27 | 0.5637 | 0.02 | 1.05 | 1.09 | 5M27C | 57 | 3.2285 | 0.0122 | 0.82 | 0.81 | 8O57C |
| 28 | -0.8028 | 0.021 | 0.97 | 1.02 | 5M28C | 58 | 2.8076 | 0.013 | 0.95 | 0.95 | 8O58C |
| 29 | -0.5954 | 0.02 | 0.96 | 0.97 | 5M29C | 59 | 0.1742 | 0.015 | 0.96 | 1.01 | 8O59C |
| 30 | -0.5803 | 0.02 | 1.04 | 1.18 | 5M30C |  |  |  |  |  |  |

Table 6

*Comparison of B-values and B-value Differences of Vertical Linking Items*

| Item ID | G3 | G4-G3 | G4 | G5-G4 | G5 | G6-G5 | G6 | G7-G6 | G7 | G8-G7 | G8 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3M04C | -1.1806 | 0.09 | -1.0883 | | | | | | | | |
| 3O15C | 0.1736 | 0.02 | 0.1897 | | | | | | | | |
| 3M05C | -1.8337 | -0.03 | -1.8663 | | | | | | | | |
| 3M02C | -1.0851 | *-0.80* | -1.8818 | | | | | | | | |
| 3O16C | 0.3078 | 0.06 | 0.3628 | | | | | | | | |
| 4O24C | 1.9522 | -0.07 | 1.8785 | *-0.40* | 1.4754 | | | | | | |
| 4M20C | -1.6195 | *-0.40* | -2.0164 | 0.03 | -1.9913 | | | | | | |
| 4M18C | 0.1823 | 0.21 | 0.3926 | -0.14 | 0.25 | | | | | | |
| 4M19C | 0.1585 | 0.00 | 0.1604 | *0.40* | 0.5583 | | | | | | |
| 5M28C | | | -0.9968 | 0.21 | -0.7911 | 0.08 | -0.7125 | | | | |
| 5M29C | | | -0.355 | -0.25 | -0.6093 | -0.12 | -0.7255 | | | | |
| 5M30C | | | -0.6954 | 0.11 | -0.5885 | 0.24 | -0.346 | | | | |
| 5M25C | | | -1.5886 | -0.23 | -1.8215 | 0.26 | -1.5584 | | | | |
| 5M26C | | | 0.6338 | -0.12 | 0.5101 | -0.20 | 0.311 | | | | |
| 5O34C | | | 2.5181 | *-0.43* | 2.0907 | *-0.34* | 1.7504 | | | | |
| 6M38C | | | | | 0.6567 | 0.02 | 0.6787 | -0.08 | 0.6029 | | |
| 6O43C | | | | | 0.0559 | 0.02 | 0.0712 | -0.03 | 0.0454 | | |
| 6M36C | | | | | 0.5018 | -0.20 | 0.3057 | 0.12 | 0.4271 | | |
| 6M39C | | | | | 1.0136 | 0.09 | 1.1032 | -0.06 | 1.0432 | | |
| 7M47C | | | | | | | 1.9168 | 0.23 | 2.1476 | 0.05 | 2.1953 |
| 7M46C | | | | | | | 0.8065 | -0.23 | 0.5744 | 0.22 | 0.7985 |
| 7M44C | | | | | | | 0.5084 | *-0.36* | 0.1526 | -0.23 | -0.073 |
| 7M48C | | | | | | | | | -0.082 | 0.11 | 0.0323 |
| 7M49C | | | | | | | | | 0.6156 | 0.00 | 0.6144 |
| 8M54C | | | | | | | | | 1.0956 | -0.01 | 1.088 |
| 8O57C | | | | | | | | | 3.2239 | 0.00 | 3.2208 |
| 8M51C | | | | | | | | | 0.9104 | *0.31* | 1.2205 |
| 8M53C | | | | | | | | | -0.3563 | -0.01 | -0.3661 |
| 8M55C | | | | | | | | | 1.4096 | 0.01 | 1.4149 |

Appendix B



```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    6         .####  +
                  .  |
                     |
                     |
                     |
                  .  |
    5             .  +
                  .  |
                     |
                     |
                  .  |
                  .  |
    4             .  +
                  .  |
                  .  |
                 .#  |
                  .  |
                  . T|  8O57C
    3         .####  +
                .##  |  8O58C
                 .#  |
             .#####  |
               .###  |T
               .###  |  5O34C   7M47C
    2           .##  +
            .###### S|  4O24C
             .#####  |
      .############  |
         .#########  |  8M55C
             .#####  |S 6M39C   8M51C   8M54C
    1   .##########  +
        .##########  |  6M40C   6M41C
          .########  |  5M32C   5M33C   6M38C   7M45C   7M49C   7M50C   8M52C
        .########## M|  5M26C   5M27C
         .#########  |  3O16C   4M18C   6M35C   6M36C
      .############  |  3M03C   3O15C   4M19C   5M31C   7M44C   8O59C
    0       .######  +M 3M11C   6O43C   7M48C
          .########  |  4M22C   6M42C   7M46C
        .##########  |  3M09C   3O17C   4M23C   8M53C
           .#######  |  3M14C   5M30C
               .###  |  3M08C   5M29C
               .### S|  3M01C   5M28C
   -1       .######  +  3M13C
              .####  |S 3M02C   3M04C
                .##  |  3M06C   8M56C
               .###  |  3M07C   3M12C
                 .#  |  6M37C
                .##  |  3M05C   3M10C   5M25C
   -2           .##  +  4M20C
                  . T|
                  .  |T
                 .#  |
                  .  |
                  .  |  4M21C
   -3             .  +
                     |
                  .  |
                  .  |
                     |
                     |
   -4            .#  +
               <less>|<freq>
 EACH "#" IS 380: EACH "." IS 1 TO 379
```

*Figure 1*. The Wright map from concurrent calibration.