



“Research Article”

Using the 2006 PISA Questionnaire to Evaluate the Measure of Educational Resources: A Rasch Measurement Approach

Ruixue Liu^{*1}, Letao Sun¹, Jing Yuan¹, Kelly Bradley¹

¹University of Kentucky, United States

Abstract

School educational resources are key when studying school improvement due to their influence on learning outcomes. Because of this, careful attention should be given to the way educational resources are operationalized and measured. Using the 2006 PISA American sample containing 166 schools, this study aims to validate the 13-item PISA School Educational Resource Scale with Rasch analysis. Winsteps software was used in the analysis and results were used to evaluate how well the instrument measured the construct of school educational resource. Findings revealed that the PISA 2006 data gave an overall indication of good fit to the model, despite the instrument not separating respondents well. In regards to the quality of the scale, the majority of items perform consistently with the model. However, for schools above the average educational resource threshold, it appears there is a need for more items to discriminate the situation.

Article Info

Received
February 15, 2017

Revised
May 27, 2017

Accepted
June 01, 2017

Keywords

School educational resource,
PISA,
Rasch rating scale model,

1. INTRODUCTION

According to Hanushek (1997), school educational resource was operationalized as the combination of the real resources of the classroom (e.g. teacher education, teacher experience, and teacher-pupil ratios), financial aggregates of resources (e.g. expenditure per student and teacher salary), and estimates of other resources in school (e.g. specific teacher characteristics, administrative inputs, and facilities). School educational resource plays a critical role in attaining educational objectives and create equal opportunities for students (Savasci & Tomul, 2013). With the Every Student Succeeds Act (ESSA), the federal government has become more deeply involved in seeking to improve student achievement. With the emphasis on the development of

*Corresponding Author E-mail: liuruixue2046@hotmail.com

student achievement, educational leaders and policymakers should make effective decisions on allocating school educational resource to help school meet student learning objective. To make these decisions, educational leaders and policymakers need reliable evidence of the effects of specific educational resources on student achievement (Sala, 2014).

This study applied the Rasch rating scale model to assess the quality of the School Educational Resource Scale, an instrument used to evaluate school educational resources in Program for International Student Assessment (PISA) 2006. Specifically, the aim of the study is to provide an overall assessment of the psychometric properties of this instrument. Findings may lead to a more accurate measure of school educational resources.

1.1. School Effectiveness Research

Studies of school educational resources have been embedded in school effectiveness research (Murnane, 1981; Schneider, 1985; Ma, 2001; Konstantopoulos, 2006; Stanco, 2012).

An effective school has been defined in different ways (Johnson, 2008). For example, Lezotte (2001) claimed that an effective school should provide “(1) instructional leadership, (2) clear vision and mission, (3) safe and orderly environment, (4) high expectations for student’s achievement, (5) continuous assessment of student achievement, (6) opportunity and time on task and (7) positive home-school relations” (p.4). Some researchers have focused on academic achievement of the students (e.g., MacNeil, Prater, & Busch, 2009; Koth, Bradshaw, & Leaf, 2008), while other researchers concentrated on differences in attitudes and behavior of the students (e.g., Elliot, Cornell, Gregory, & Fan, 2010; Way, Reddy, & Rhodes, 2007).

The following effective school definition was adopted by the Organization for Economic Cooperation and Development ([OECD], 1994) with a global approach: “*An effective school promotes the progress of its students in a broad range of intellectual, social, and emotional outcomes, while considering socio-economic status, family background and prior learning*” (p.1).

School effectiveness studies covered three generations over the past 50 years (Fan, 2013). The first generation of school effectiveness research started about 50 years ago with the publication of Coleman and his colleagues’ (1966) research on the quality of schooling in the United States. This study, known as *The Coleman Report*, has been regarded as the first large-scale study of school effectiveness and considered as the major impetus for development of research in this field (Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002). In this study, the results of standardized test of ability and achievement for a total of 645,000 students from more than 4,000 schools were collected and analyzed to explore whether the schools had a measurable impact on student achievement. Coleman et al. concluded that schools have relatively little impact on student achievement compared to the socioeconomic background and started an ongoing debate.

A group of noteworthy school effectiveness studies in the mid-1980s, including the *School Matters* in London (Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988) and *Louisiana School Effectiveness Study* (LSES) (Teddlie & Stringfield, 1993), were considered the second generation of school effectiveness studies. In the study of *School Matters*, Mortimore et al. (1988) aimed to examine the size of school effect, differentiate school effectiveness, and identify factors to develop school effectiveness. Two thousand children, randomly selected from 50 primary schools participated in this study over the course of four years. The LSES was a longitudinal study conducted from 1980 to 1992, utilized both quantitative and qualitative methods (Teddlie et al. 1993) in the United States. This was a longitudinal study from 1980 to 1992 which utilized both

quantitative and qualitative methods to analyze data at the school and classroom levels. Several factors to promote effectiveness of middle school with low SES were discovered and discussed, including the enhancement of educational expectations; principal leadership style; usage of external reward structures; the emphasis on school curriculum; parental involvement; and the experience level of teachers.

In the third decade, the school effectiveness research shifted toward a globalization in the field (Teddlie & Reynolds, 2000). The majority of school effectiveness studies have been conducted in the western countries such as the United States, the United Kingdom, the Netherlands, Australia and Canada. As Teddlie (2004) called attention to and is still the case today, it is necessary to also study under-represented areas of the world to enrich the knowledge base of this field and to make comparisons with the existing research.

1.2. School Educational Resources

Many studies have researched the question of whether the level, or amount, of school educational resources influenced student outcomes of learning. Unfortunately, it has proven difficult to determine the relationship between school educational resources and student achievement outcomes (Sala, 2014). According to Hanushek (1997), evidence was not found to support a strong or consistent relation between school educational resources and student achievement. This finding has received considerable attention and acceptance by individuals in the academic, legal, and public policy arenas. Others have challenged this position and results from other studies provide counter evidence. Knoepfel, Verstegen, and Rinehart (2007) found that average school wealth has positive effects on student achievement. Moreover, Jacob and Ludwig (2008) showed that increased funding used in early childhood education, class size reduction, and salary lead to improved student outcomes. Vandiver (2011) indicated that quality and educational adequacy of educational facilities were statistically significantly correlated with student performance.

According to the Organisation for Economic Co-operation and Development (2010), effective school systems require the right combination of qualified personnel, adequate educational resources, facilities, and motivated students ready to learn; in addition, factors including class and school size, the quality of teaching materials, perceived staff shortages, and teacher quality are frequently associated with student performance. Most noticeably, school educational resources are the most important set of mediators through which the socio-economic background of students and schools affects performance.

The mixed findings on the effectiveness of school educational resource on academic achievement may partly due to instruments with an inadequate quality. Thus, it is necessary to develop a more reliable and valid instrument to measure school educational resources. The Rasch model, as a powerful approach to investigate psychometric properties, was conducted in this study. The following section will provide a brief introduction of Rasch model.

1.3. Rasch Model

According to Wright and Linacre (1989), the arithmetical property of interval scales is fundamental to any meaningful measurement. Traditional analytical techniques usually anchor on True Score Theory, and the raw data are not interval data. Thus, the data only indicate ordering without any proportional meaning (Yan & Mok, 2012). According to Waugh and Chapman (2005), one cannot make valid inferences from the measures that are initially set up for True Score Theory.

The aforementioned issue can be overcome by analyzing the data via the Rasch model. The Rasch model, introduced by Georg Rasch (1960), can generate a comprehensive picture of the association between observed item responses on a scale and persons' levels on a latent variable. The Rasch model is the simplest of the Item Response Theory (IRT) models, having a single parameter for the person or entity and a single parameter corresponding to each category of an item. An application of the Rasch model is appropriate any time a researcher wishes to use the total score on an assessment or questionnaire to make inferences about an individual's ability or level of a latent trait inherent in that individual (Bond & Fox, 2001).

Since the Rasch model arises from the requirement that comparisons among person and items are invariant across samples, it is appropriate when the total score on a test or questionnaire is used to make inferences. Although Classical Test Theory (CTT) also uses the total score to characterize each person, the total score is used as the relevant statistics without paying enough attention on the anomalies in the items or persons answering them. These anomalies can be explained by the Rasch model which can provide a more informative score. The objective of Rasch measurement is similar with the construction of a ruler, establishing the correct measure (Andrich & Luo, 2003).

The Rasch model is a methodological tool that can be used to analyze data, especially when dealing with latent traits such as attitudes or perceptions. It allows observations of respondents and items to be connected in a way that indicates the occurrence of a certain response as probability rather than certainty and maintains order in that the probability of providing a certain response defines an order of respondents and items. In other words, a person endorsing an extreme statement, or answering a difficult item, should also endorse all less extreme statements, or answer correctly the less difficult items (Wright & Masters, 1982). A rating scale is a set of categories designed to elicit information about a quantitative or a qualitative attribute. In the social sciences, a common example is the use of a Likert scale in which a person selects the number which they consider to reflect the perceived quality of a product (Andrich, 1978). In the current study, the rating scale model was used, as it is appropriate for the analysis of survey data. The formula is:

$$\ln \left(\frac{p_{nij}}{p_{ni(j-1)}} \right) = B_n - D_i - F_j \quad 1$$

In Equation 1, p_{nij} = the probability that person n encountering item i is observed in category j , B_n = the "ability" measure of person n , D_i = the "difficulty" measure of item i , (the point where the highest and lowest categories of the item are equally probable), F_j = the "calibration" measure of category j relative to category $j-1$ (Rasch-Andrich threshold located at the point of equal probability of categories $j-1$ and j); and no constraints are placed on the possible values of F_j . Winsteps measurement software was used to perform the Rasch analysis (Linacre, 2009).

2. METHOD

2.1. Data Source

The primary database used in this research is constructed from the Program for International Student Assessment (PISA) conducted in 2006. According to Organization for Economic Cooperation and Development (OECD) (2001), PISA is the most comprehensive and rigorous international assessment on 15-year-old student performance in reading, science, and mathematics.

Every three years, data is collected on the student, family and institutional factors that is used to analyze differences in performance. PISA examines how well students are prepared to meet the challenges of the future and how well students are prepared for life in a larger context, rather than how well they master particular curricula. In 2006, PISA included information on nearly 400,000 students from 57 countries. The database included student performance in reading, science, and mathematics. In addition, data from the parents and school principals of participating schools were also included.

The data for this study is derived from the United States sample in the 2006 PISA study conducted by OECD. Data were downloaded from the OECD website. SPSS 22.0 program was used to manage and clean the data. The sample contains 166 persons (high school principals). Eleven persons who failed to complete this survey were excluded from the Rasch analysis. Therefore, there were 155 persons measured on the 13 items for this study.

2.2. Instrument

The entire set of items used in this scale is derived from the school questionnaire of PISA 2006. The index of school educational resource aims to measure principals' perceptions of potential factors hindering instruction at schools through the 13-item scale (e.g., a lack of qualified science teachers; shortage or inadequacy of science laboratory equipment; shortage or inadequacy of computer software for instruction; Shortage or inadequacy of audio-visual resources). A four point Likert-type scale was used (not at all = 1, very little = 2, to some extent = 3, a lot = 4). As all items were inverted for scaling, higher values on this index indicate more school educational resources. The detailed items can be found in Table 1.

Table 1. Items of School Educational Resource Assessment

Question	Items	Responses
Is your school's capacity to provide instruction hindered by any of the following?	1. A lack of qualified science teachers	1 - Not at all
	2. A lack of qualified mathematics teachers	2 - Very little
	3. A lack of qualified (test language) teachers	3 - To some extent
	4. A lack of teachers of other subjects	4 - A lot
	5. A lack of laboratory technicians	
	6. A lack of other support personnel	
	7. Shortage or inadequacy of science laboratory equipment	
	8. Shortage or inadequacy of instructional materials	
	9. Shortage or inadequacy of computers for instruction	
	10. Lack or inadequacy of Internet connectivity	
	11. Shortage or inadequacy of computer software for instruction	
	12. Shortage or inadequacy of library materials	
	13. Shortage or inadequacy of audio-visual resources	

3. RESULTS

Dimensionality Analysis: The Rasch principal components analysis of residuals was carried out to assess the dimensionality of the constructed scale. The eigenvalue of the first contrast was 3.3, indicating it has the strength of about three items (3.3 rounded to 3, out of 13). It is larger than the strength of two items (an eigenvalue of 2), the smallest amount that could be considered a dimension. Meanwhile, the eigenvalue of second contrast is 1.8. Thus the assumption of unidimensionality holds, and is not violated, in this study.

Reliability and Separation: Both reliability and separation statistics can be considered at the person and item level. Person reliability is analogous to Cronbach's alpha reliability in True Score Theory while item reliability has no traditional equivalent. Low values for item reliability indicate a narrow range of item measures, or a small sample. Person separation is used to classify people, and item separation is used to verify the item hierarchy (Linacre, 2009). The reliability and separation statistics can be found in Table 2. Person reliability was computed to be 0.76, and item reliability was 0.90. Person separation was 1.76, and item separation was 3.07.

Table 2. Model Fit Statistics

	Measure	Infit ZSTD	Outfit ZSTD
Principals (Reliability= .76; Real RMSE=.70)			
<i>M</i>	42.20	-.10	.00
<i>SD</i>	6.60	1.30	1.30
Items (Reliability= .90; Real RMSE=.13)			
<i>M</i>	502.80	-.10	-.10
<i>SD</i>	27.70	1.90	1.60

Model Fit Statistics: ZSTD is a *t*-test of the hypothesis "Do the data fit the model (perfectly)?" They are reported as z-scores. Besides, they show the improbability of the data, if the data actually fits the model. Zero are their expected values. Less than 0 indicates too predictable. More than 0 indicates lack of predictability. Generally, if the ZSTD were within the range of -1.9 to 1.9, the instrument indicates a reasonable predictability (Linacre, 2002). Table 2 showed that both the infit and outfit ZSTD could meet this requirement.

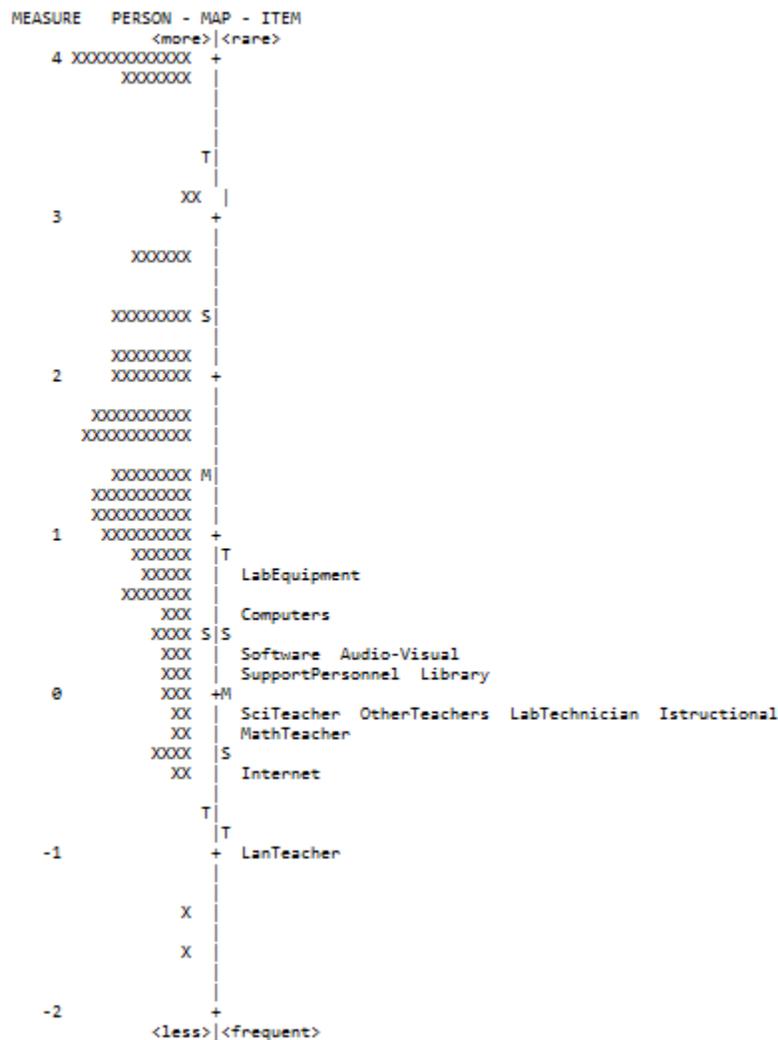
Item Infit and Outfit: There are two types of item fit statistics in the Rasch analysis. Item outfit statistics are influenced by unexpected responses to items, for example, when a person of low ability gets a very difficult item correct. Infit statistics are influenced by an unexpected pattern of responses near a person's ability estimate, for example, when a person gets the item near the person's ability estimate incorrect.

Table 3 shows the item misfit statistics, which reveals several misfitting items. For instance, Item 2 (A lack of qualified mathematics teachers) has the maximum infit indices (ZSTD = 3.50), which exceed the upper bound of criteria range of infit ZSTD ($3.50 > 2$), and Item 11 (Shortage of inadequacy of computer software for instruction) has the minimum infit indices (ZSTD = -3.20), which exceed the lower bound of criteria range of infit ZSTD ($-3.2 < -2$). In addition, Item 11 also has the minimum outfit indices (ZSTD = -2.60) that exceed the lower bound of criteria range of outfit ZSTD ($-2.60 < -2$) (see Table 3).

Table 3. Item Statistics

Items	Measure	Infit	Outfit
		ZSTD	ZSTD
2	.23	3.50	2.80
1	.07	2.20	1.60
5	.12	1.90	1.70
3	.94	.70	.90
6	-.18	1.00	1.30
9	-.50	.90	1.00
7	-.78	-.40	-.20
4	.17	-.50	-.040
10	.44	-1.10	-1.10
13	-.23	-1.80	-1.80
12	-.14	-1.90	-1.70
8	.12	-2.40	-2.20
11	-.29	-3.20	-2.60

Figure 1. Item-person map for school resource items



Item and Person Map: Figure 1 shows the item-person map, which provides distribution for both item difficulty and person ability estimates on a single line of logit scale to facilitate the graphical representation of the relationships. This map displays the person measure and item measure on the same scale. The ability estimates are shown on the left side and the item difficulty locations are shown on the right. Person ability and item difficulty increase as one moves towards the top of the figure (Linacre, 2009). Overall, this map shows that the majority of person ability distribution falls outside of the range of the item difficulty distribution. Persons' ability scoring around 0 logits are found to be well measured by the items, and all item difficulty estimates are clustered around 0 logits. However, the ability distribution is higher overall than the difficulty distribution, which indicates that persons with higher ability are not accurately, or maybe fully, measured by the items.

4. DISCUSSION

The item separation and reliability statistics showed that the person sample is large enough to confirm the item difficulty hierarchy (construct validity) of the instrument. However, low person separation (less than 2) and person reliability (less than 0.8) implied that the instrument may not be sensitive enough to distinguish between high and low performers. Adding more items could be a solution to the issue. Meanwhile, the analysis of misfit reveals some potentially misfitting items on the school educational resource scale, suggesting revision may be needed. The item-person map reveals that persons with higher ability are not accurately measured by the items.

The central focus of school effectiveness research concerns the idea that "*schools matter, that schools do have major effects upon children's development and that, to put it simply, schools do make a difference*" (Reynolds & Creemers, 1990, p. 1). Moreover, as mentioned earlier, many studies have examined the question of whether the level, or amount, of school educational resources influences the level, or outcomes, of student learning. Some studies indicate that school educational resources do not have an effect on academic achievement of students (Hanushek, 1997; Hanushek & Luque, 2003). On the other hand, some studies say the exact opposite (Card & Krueger, 1996; Greenwald et al., 1996). This debate leads to researchers seeking instruments to measure school educational resources. With so many instruments, some of them may not be high quality measures, illustrating poor quality in terms of the reliability and validity. Instruments with low reliability may produce different results under comparable, consistent conditions. Validity can help determine what types of assessments to use and make sure whether a method can truly measure the idea or construct in question. Because of this, careful attention should be given to the way educational resource is operationalized and measured and developing a more reliable and valid instrument to measure school educational resources may be the most important part of conducting a high quality research study in this area.

Above all, using a powerful technique to evaluate the psychometric properties of an instrument is important. The current study evaluated how well the instrument measured the construct of school educational resources by analyzing the constructed scale. A good Likert-type scale is grounded in sufficient items with a varying degrees of difficulty to evaluate a range of abilities held by the persons. Utilizing the Rasch model to analyze survey research data will result in more sound measures and more meaningful results (Bond et al., 2001). For example, the Rasch model produces estimates of the latent trait displayed by each subject ("person measure") and the trait to respond in a certain way to each item ("item measure"). The Rasch model also provides item fit statistics that indicate whether the individual item is contributing to the measurement of

the latent trait (Bond et al., 2001). Furthermore, the Rasch model software (e.g., Winsteps) can provide indices and visual displays that help examine whether items and persons spread sufficiently along the continuum of the measure (Linacre, 2009). This enables survey researchers to visualize if and where additional items are necessary to cover the entire dimension of the construct. Above all, researchers and practitioners in testing and measurement should be aware of the advantages of using Rasch analysis.

5. CONCLUSIONS

In the current study, the Rasch analysis' results provide a more detailed and comprehensive display of how school principals perceive potential factors hindering instruction at their schools. These results could be disseminated to provide PISA administrators with useful information to make more informed decisions regarding survey administration methods and the interpretation and comparability of the impending results. By using the same framework, the Rasch analysis can be used to examine other school context and climate variables (e.g., teacher effectiveness, classroom practice, and principal leadership) in the school effectiveness research, large-scale assessment, and international comparative studies.

The results of this study, which employed the Rasch measurement model to analyze the PISA 2006 data, give an overall indication of good fit to the model. There were two major weaknesses of the instrument brought to light through this analysis. On the one hand, the item-person map and the statistics of person separation and reliability indicate that there are not enough items to discriminate the situation of school educational resources for schools that are above the average. Even so, this might not matter, as those above average might have reached a successful plateau. On the other hand, some misfitting items were discovered by the analysis of misfit, and they are suggested to be revised in the future research.

The alignment between accountability policies and school finance policies to better serve student learning goals has been emphasized by educational researchers (Superfine, 2009). Findings of this study can contribute to the future research on the effects of school educational resources on student academic achievement. To this end, educational policymakers will have reliable evidence of school educational resources to inform resource allocation practices to meet the demands of educational adequacy.

6. REFERENCES

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 357-74.
- Andrich, D. & Luo, G. (2003). Conditional pairwise estimation in the Rasch model for ordered response Categories using principal components. *Journal of Applied Measurement*, *4*(3), 205-221.
- Ammermueller, A., Heijke, H., Woessmann, L. (2005). Schooling quality in Eastern Europe: Educational production during transition. *Economical Educational Review*, *24* (5), 579-599.
- Bond, T., & Fox, C. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Card, D., & Krueger, A. (1996). School resources and student outcomes: An overview of the literature and new evidence from North and South Carolina. *Journal of Economic Perspectives*. doi: jep.10.4.31

- Coleman, J. S., Hoffer, C., & York, R. (1966). *The equality of educational opportunity study*. Washington, DC: United States Department of Health, Education, and Welfare.
- Dodson, C. K. (2005). *The Relationship between School Effectiveness and Teachers' Job Satisfaction in North Mississippi Schools*. Unpublished Doctoral Dissertation, Mississippi University, Oxford.
- Eliot, M., Cornell, D., Gregory, A., & Fan, X. (2010). Supportive school climate and student willingness to seek help for bullying and threats of violence. *Journal of School Psychology, 48*, 533-553. doi:10.1016/j.jsp.2010.07.001
- Fan, M. (2013). *Stability of academic performance across science subjects among Chinese students* (Unpublished master's theses). University of Kentucky, Lexington, KY.
- Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141-164. doi:10.2307/1164207
- Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review, 22*, 481-502. doi: 10.1016/S0272-7757(03)00038-4
- Knoeppel, R. C., Verstegen, D. A., & Rinehart, J. S. (2007). What is the relationship between resources and student achievement: A canonical analysis. *Journal of Education Finance, 33*(2), 183-202.
- Jacob, B. & Ludwig, J. (2008). *Improving Educational Outcomes for Poor Children*. Cambridge, MA: National Bureau of Economic Research.
- Johnson, A. D. (2008). *The relationships among middle school student and staff perceptions of school effectiveness and student achievement*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.
- Koth, C., Bradshaw, C., & Leaf, P. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. *Journal of Educational Psychology, 100*(1), 96-104. doi: 0022-0663.100.1.96
- Lezotte, L. W. (2001). *Revolutionary and Evolutionary: The Effective Schools Movement*. Retrieved from <http://www.edutopia.org/pdfs/edutopia.org-closing-achievement-gap-lezotte-article.pdf>.
- Linacre, J. M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions, 16*(2), 878.
- Linacre, J. M. (2009). A user's guide to Winsteps, Ministep, Rasch-model computer programs: Program manual 3.72.3. Retrieved from <http://www.winsteps.com/a/winsteps-manual.pdf>.
- Ma, X. (2001). Stability of school academic performance across subject areas. *Journal of Educational Measurement, 38*(1), 1-18.
- MacNeil, A., Prater, D., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of Leadership in Education, 12*(1), 73-84. doi: 10.1080/13603120701576241
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters*. Berkeley, CA: The University of California Press.
- Murnane, R. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record, 83*(1), 19-35.
- Organization for Economic Cooperation and Development. (1994). *Making education count: Developing and using international indicators*. Paris: Author.

- Organization for Economic Cooperation and Development (2010). *PISA 2009 Results: What Makes a School Successful? – Resources, Policies and Practices (Volume IV)*. Paris: Organization for Economic Cooperation and Development.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Reid, K., Hopkins, D., & Holly, P. (1987). *Towards the effective school*. Oxford: Blackwell.
- Reynolds, D & Creemers, B. (1990). School effectiveness and school improvement: A mission statement, *School Effectiveness & School Improvement*, 1(1): 1-3. doi: 10.1080/0924345900010101
- Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (2002). *World class school: International perspectives on school effectiveness*. London: Routledge Farmer.
- Savasci, H. & Tomul, E. (2013). The relationship between educational resources of school and academic achievement. *International Education Studies*, 6(4), 114-123. doi:10.5539/ies.v6n4p114
- Sala, M. (2014). *Examining the effects of school-level variables on elementary school students' academic achievement: The use of structural equation modeling* (Unpublished doctoral dissertation). Clemson University, Clemson, SC.
- Smith, R. M. (1996). Polytomous mean-square fit statistics. *Rasch Measurement Transactions*, 10, 516-517.
- Schneider, B. (1985). Further evidence of school effects. *Journal of Educational Research*, 78(6), 351-356.
- Stanco, G. (2012). *Using TIMSS 2007 data to examine STEM school effectiveness factors in an international context*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.
- Superfine, B. M. (2009). Deciding who decides questions at the intersection of school finance reform litigation and standards-based accountability policies. *Educational Policy* 23(3), 480-514. doi: 10.1177/0895904808314712
- Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: lessons learned from a 10-year study of school effects*. New York: Teachers College Press.
- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.
- Teddlie, C. (2004). Getting schools working in South Africa: A valuable addition to the SESI field. *School Effectiveness and School Improvement*, 15(2), 227-240.
- Vandiver, B. (2011). *The impact of school facilities on the learning environment* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing (UMI No. 3439537)
- Waugh, R. & Chapman, E., (2005). An analysis of dimensionality using factor analysis (true score theory) and Rasch measurement: what is the difference? Which method is better? *Journal of Applied Measurement*, 6(1), 80-99.
- Way, N., Reddy, R., Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: Associations with trajectories of psychological and behavioral adjustment. *American Journal of Community Psychology*, 40, 194–213.
- Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal: Measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857-860.

- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.
- Yan, Z. & Mok, M. (2012). Validating the coping scale for Chinese athletes using multidimensional Rasch analysis. *Psychology of Sport and Exercise, 13*, 271-279. doi:10.1016/j.psychsport.2011.11.013.