A school-based evaluation of reading gains and the use of a value-added grid for making interpretations of assessment results

Harrison Tse
Hong Kong Institute of Education

Stephen Albone
Durham University

Abstract

Background: In line with the increase of using value-added assessment information to evaluate learning gain made by learners in different parts of the world, the Education Bureau in Hong Kong has recently introduced a system that provides value-added information of secondary school students. However, not many teachers have a clear knowledge and understanding about value-added assessment.

Aims: This paper provides a brief review of the meaning and the importance of value-added assessment in education. Then it provides an example of how to find out the academic attainment that one can reasonably expect from individual students and how to produce two types of school-based value-added estimations with reference to their own prior attainment or development ability measures. The paper also evaluates the use of descriptions in the value-added grid for making interpretations of the value-added assessment results.

Sample: 60 schoolchildren in a primary school in England.

Research method: Linear regression models.

Results: The findings indicated that the descriptions in the grid were applicable for interpreting the assessment results, but attention had to be paid when making speculations about the positions of individual schoolchildren's past attainment in reading.

Conclusion: This paper functions as a simplified example and a timely resource for teacher learning and development concerning value-added assessment. It also reflects the importance for teachers and school administrators to equip themselves with a good understanding of the nature, potential value and meanings of value-added estimations.

Keywords*:* value-added estimation, assessment of reading, provision of feedback

閱讀增益的校本評鑑和使用增值網格來解釋評估結果

謝加謙〔香港〕， S. Albone〔英國〕

摘要

背景：隨着世界不同地區更廣泛地使用增值評估資料來評鑑學習者的學業增益，香港教育局近年亦引入了一套提供中學生增值資料的系統。然而，許多老師仍對增值評估缺乏認知和了解。

目的：本文為增值評估在教育的意義及其重要性提供一篇簡要的回顧; 然後提供一個尋找對個別學生學術成就作出合理期望的例子，並參考他們自己較早前的學術成就或自己已發展的能力，而產生出兩種以學校為本的增值估計。本文亦對使用增值網格內的描述來解釋增值評估結果作評鑑。

研究對象：60位在英國一間小學就讀的學童。

研究方法：直線回歸統計模型。

研究結果：結果顯示了增值網格內的文字描述，可用作解釋這校本增值評估結果，但須要留意它對個別學童過往閱讀水平所作出的推測。

總結：本文提供了一個增值評估的簡化例子和相關的教師進修與發展的適時資料。它亦給老師和學校行政管理人員反映一個重要的訊息，他們需要裝備自己對增值估計的本質、潛在價值和意義的了解。

關鍵詞：教育增值估計、閱讀評估、提供回饋。

Background

Movements in educational assessment and evaluation over the last thirty years have shown that value-added information is a useful indicator reflecting learning gains made by schoolchildren. There has been growing interest in the use of the information as a form of evidence in the evaluations of educational effectiveness, school effectiveness or teacher effectiveness in different places around the world (e.g. Kelly and Monczunski, 2007; Tse, 2004; Olson, 2004; Lee and Landauer-Menchik, 2002; Tymms, 1999; Saunders, 1999; Sanders and Horn, 1998; Tymms, 1997; Fitz-Gibbon, 1997; Tymms and Fitz-Gibbon, 1995). In Hong Kong, the Education Bureau has formally introduced a value-added assessment system to secondary schools as an education policy of the government (EDB, 2007). It is expected that this trend of innovation in assessment would also extend to primary schools. However, it appears that not many administrators, practitioners and professionals working at primary and secondary education levels have a clear understanding of the issues outlined below:

1. What does the word "value-added" mean in education?
2. Why is value-added assessment becoming increasingly important in education?
3. How are value-added results produced?
4. How to interpret the information about student learning gains?

Given that it is impossible to address all the key issues and queries about value-added assessment with the limited space here, this paper will focus on providing, widening, strengthening, and/or consolidating readers' basic knowledge and understandings about the meaning of learning gain in the context of school-based data analysis. The first and second questions will be addressed through descriptions of the background of this paper and the review of literature about value-added assessment and learning gain in reading. It is hoped that the descriptions will foster readers' awareness of value-added assessment and update their knowledge and understanding of such a world-wide growing fashion in educational assessment and evaluation.

The third question will be addressed through an example of how to find out the reading attainments that one can reasonably expect from individual students and how to calculate two types of school-based value-added estimations in reading, with reference to their own prior attainment measure and developed ability measure respectively. Descriptions of these measures will be provided in the text below. Then, to address the fourth question, the paper will also introduce and evaluate the use of a value-added grid for making interpretations of the assessment results. In relation to the role and responsibility to use the assessment information for setting appropriate literacy learning targets and for planning and evaluation of language teaching, answers to the third question would be of interest to school administrators, language teaching advisors, educational policy makers and other educational professionals.

So far many secondary school teachers would have heard about value-added assessment as a response to the value-added system introduced by the Education Bureau [EDB], however, their attention could have been focused on student value-added information collected over time. The significance of another type of value-added information and the possibility to combine the two types of information could have been neglected or unfamiliar to them. To address the fourth question, it would also be of interest to introduce them to the value-added grid

as a tool for combining and making interpretations of the two types of value-added assessment results. Information in the grid has implications to the work as a reading support specialist, a language teacher or any other educational practitioner for identifying schoolchildren who are at-risk in reading progress, assessing and monitoring the attainment and progress in reading of individual schoolchildren.

Given that there is hardly any local example that illustrates the two types of value-added estimations, this paper has intended to make a leap forward by providing a simplified example of school-based value-added analysis of data concerning students of a primary school in the United Kingdom. It has no intention to promote a replication of the value-added assessment process without adaptations to the Hong Kong educational context, but such an example would likely become a timely reference resource for professional development of teachers in relation to the growing fashion of value-added assessment. Nevertheless, this paper also illustrates a practical evaluation of the use of a value-added grid, for making interpretations about learning gains made by the students.

Reading gains and value-added assessment

In addition to information about "attainment", there is recent literature suggesting that information about "learning gain" is a predominant area to consider (e.g. Kelly and Monczunski, 2007; Olson, 2004; Saunders, 1999) for making judgments in educational assessment and evaluation. The former refers to the educational standard that learners have reached, whereas the latter refers to the advance in attainments that they have made as a result of learning and instruction. In the perspective of an individual learner, a value-added measure is an indicator of relative advantage that a learner has gained from learning and instruction, such as the difference in attainments at the start and the end of an educational arrangement that has taken place over a period of time. The following text will further explain its nature, importance, and development.

The term "reading gain" here specifically refers to learning gain in reading. Instead of addressing different types of learning gains, the scope of this paper is on the use of value-added information in a school-based evaluation of reading gains. Reading gain is selected because it is regarded as one of the major types of academic learning gains. In addition to listening, speaking and writing, reading is one of the four language skills that people use in everyday life. To become literate, schoolchildren have to acquire the skills of reading and writing. Furthermore, the ability to read is often regarded as an important tool that helps them in learning to learn (CDC, 2001). This paper will produce two types of value-added in reading estimations and they will also be employed to evaluate the validity of the value-added grid as an assessment tool for making interpretations and feedback about the academic learning gains of the schoolchildren. In contrast, the scope of this paper will not cover issues about the possibility, existence or implication of value-added information concerning non-academic learning gains, such as value-added information about socio-emotional outcomes, or skill-related learning outcomes.

Nature and importance of value-added assessment

The term "value-added" is derived from economics. It refers to the difference between inputs and final outputs represented by the value of sales. For example, if we buy an apple for 4 dollars, then

turn it into an apple pie and sell it for 6 dollars, we can say that the added value is 2 dollars. Similar to the increase in value through production, education is a process that brings an increase in the value of human resources to society. It can be regarded as a process of equipping students with knowledge, attitudes, skills or qualifications that lead to relative advantages, such as better employment opportunities and personal qualities.

In general, all schoolchildren gain from learning and instruction as they pass through the education processes over a period of time (e.g. DCSF, 2007). It is worthwhile to consider that there is a variation in the extent of relative advantage between learners because there are various factors affecting their learning gains. This paper focuses on two types of relative advantages: advantage in terms of relative progress and advantage in terms of relative position in learning attainment. If the extent of progress of a schoolchild is larger than the average progress of other schoolchildren with similar attainment in the previous assessment, then that schoolchild has a relative advantage in learning gain after the prior attainment factor has been taken into account. If the attainment of a schoolchild is better than the average attainment of other schoolchildren with similar ability, then that schoolchild has a relative advantage indicated by the relative position in attainments among those schoolchildren after the developed ability factor has been taken into account. The measure of the former is called prior value-added estimation (PVA) and the measure of the latter is called concurrent value-added estimation (CVA). Further descriptions about them can be found in the text below.

Value-added estimation is often presented as a residual in assessment and evaluation because it is estimated as a "left over" after the specific background factor has been taken into account, such as the prior attainment factor or the developed ability factor. Therefore, the relative advantage of that schoolchild might be attributed to other relevant contextual factors, such as efforts put into learning, inputs from teaching and learning resources, and quality of instruction. On this basis, the value-added measure is becoming increasingly important because it is regarded as a fairer indicator for assessing and evaluating educational effectiveness (e.g. EDB, 2007; Sharp, 2006; Saunders, 1999 and Tymms and Henderson, 1995). As it takes account of the differences in input, value-added measures can be estimated after the influence of the specific input-related factor has been taken into consideration.

Development of value-added assessment in education

In the last three decades, there have been considerations of using value-added information to evaluate educational effectiveness of schoolchildren in different countries, such as the United States, the United Kingdom, Canada, Australia, New Zealand, Netherlands and Hong Kong.

The development of value-added assessment in various places of the United States originated from the demand for accountability at various levels of the education system. This includes the search for indicators of student progress and instructional effectiveness, and the need for evaluation of reform initiatives implemented by reading support specialists, school administrators or policy makers in education. Their value-added assessment model relies heavily on longitudinal data of academic attainments, and the meaning of value-added information is often represented by progress made by individual students

over time, such as the advance of their learning from one year to the next. Value-added estimation is often produced by tracking the individual student over time (e.g. Doran & Fleischman, 2005; Lee & Landauer-Menchik, 2002). The potential of producing and using value-added estimation with cross-sectional data is largely ignored.

It is worthwhile to consider that prior value-added estimations are used for retrospective evaluation of the progress made by students in a school or in an educational program over a period of time. Nowadays, the advance in educational evaluation makes it possible to evaluate the progress made by students on the basis of a growth model, which is an educational assessment framework to examine the development of individual students over a period of time. Both the value-added model and the growth model adopt a longitudinal approach to estimation, but the latter is focused on producing predictive estimations to be used as a benchmark of growth. In fact, some districts in the United States have set predictive value of "adequate growth" targets, and performance of the schools or the students is evaluated against specific estimations of the expected growth (ED, 2009).

In the United Kingdom, the development of value-added assessment was also closely linked to accountability purposes (Saunders, 1999). There were concerns about ways of making fairer comparisons of school performances, including those at secondary and primary levels. Researchers, academics and people in the educational evaluation field kept highlighting that the publication of raw results of school attainments could be misleading because the differences in input were not taken into account. The National Value Added project was a milestone to a new era in education when value-added measures were put on the National agenda. The types of value-added information studied in the project was not just restricted to the measure of attainment-related progress over a period of time, but it also included the measure of the advance in position of attainment when compared with other schoolchildren in the assessment cohort. We shall introduce these two types of value-added measures in the descriptions about the Performance Indicators in Primary Schools project (PIPS) later in the paper.

In Hong Kong, the EDB has proposed a statistical model of assessing students' results in senior secondary education. Official documents clearly state that a prior attainment or ability measure is likely to be the best choice in the selection of an independent variable for the statistical model (EDB, 2007). However, the value-added assessment system in Hong Kong relies heavily on the use of longitudinal data of student academic attainments. In addition to the prior value-added measure, it would be of significance to introduce a concurrent value-added measure through this paper. The authors believed that a study at the primary education level could be particularly useful because it is such an important stage of child development that requires educational assessment, evaluation, management and support.

With reference to the principle of "no child should be left behind" in the United States and a similar view presented in the website of the Department of Children, Schools and Families (DCSF) in the United Kingdom, this paper assumes that every schoolchild is able to gain a relative advantage from learning and instruction after going through the educational processes over a period of time (DCSF, 2007; Kelly and Monczunski, 2007; Olson, 2004). The value-added estimation in this paper may appear as a positive or negative value.

A positive value-added estimation means that the relative advantage gained by the schoolchild is better than expected. On the contrary, a negative value-added estimation means that the relative advantage gained by the schoolchild is not as well as expected. This raises an important question, "How to set a reasonable expectation?"

So far, research findings have suggested that the value of the expected academic attainment and value-added estimation can be calculated on the basis of a series of statistical data that models the academic attainment that one might reasonably expect, when prior attainment is taken into account. In some value-added assessment studies (e.g. Olson, 2004; Saunders, 1999), prior attainment is found to be the best "predictor" (i.e. also named as "independent variable") of academic attainment. Furthermore, findings of some value-added assessment studies also indicate that the value of the expected academic attainment and value-added estimation can be calculated on the basis of a series of statistical data that models academic attainment that one might reasonably expect, when developed ability is taken into account (e.g. Tse, 2004; Fitz-Gibbon, 1997). These findings provide justification to use developed ability (DA) and prior attainment (PA) as the independent variables in the first and second regression models of this paper respectively. In terms of the time factor, the data collection of the former is cross-sectional while the data collection of the latter is longitudinal in nature. In each of the two models, the dependent variable is the reading attainment (RA) of the individual schoolchild and the outcome of statistical estimation is the expected attainment in reading (E_RA).

It has been suggested that the two types of value-added estimations correlate highly with one another (e.g. see Tymms & Albone, 2002 for detail). The PIPS thinks that it is good to combine the two types of value-added information for a better and holistic picture of assessment (PIPS, 1999 and 2008). Furthermore, the PIPS has introduced the use of a "value-added grid", which is a table containing descriptions to be used for making interpretations about value-added assessment in relation to various combinations of the two types of value-added information, namely CVA and PVA. As presented in Figure 1, there are vertical and horizontal lines to divide the grid into 9 squares, with descriptions in each of the squares. The two horizontal lines are the 1st and the 3rd quartiles of the CVA. That means, with reference to the measure, 25% of the student participants in the PIPS belong to the 3 squares in the top row of the grid (i.e. + or + +), 50% of the participants belong to squares in the middle row and the remaining 25% of the participants belong to those in the bottom row (i.e. - or --). The two vertical lines are the 1st and the 3rd quartiles of the PVA, while 25% of the student participants in the PIPS belong to the 3 squares in the left-hand column of the grid, 50% of the participants belong to squares in the middle column and the remaining 25% of the participants belong to those in the right-hand column.

The grid was developed as an assessment feedback tool for categorizing students who participated in the PIPS according to the characteristics of learning gains, but the validity of the statements in the grid has not been formally investigated. The most relevant information about its use was reported in a PIPS Newsletter (PIPS, 1999), which stated that the percentage of participants who belong to the top left-hand square or the bottom right-hand square of the grid was less than 1%. Therefore, it may be reasonable to expect that the usage of the

descriptions in the two squares is rather infrequent.

In summary, the combination of the two types of school-based value-added in reading estimations and the application of the results to evaluate the validity of the value-added grid as an assessment feedback tool for making interpretations about value-added assessment are research gaps to be filled in this paper.

Figure 1: The value-added grid (Extracted from: PIPS, 1999)

		Value-added Grid		
Concurrent Value-added	+ or ++	This child is doing better than expected, however they are no longer as far ahead as they used to be.	This child is doing better than expected and this may have been a consistent characteristic over time.	This child was probably on track before. They have made excellent progress and have now moved further ahead.
	0	This child is doing as well as expected. However they have moved from a position where they were ahead of similar children.	This child is on track and this is probably a consistent characteristic over time.	This child was probably underachieving before, however they have made excellent progress and are now on track.
	- or –	This child was probably on track before but has fallen behind and is now underachieving.	This child is underachieving and this may have been a consistent characteristic over time.	This child was probably underachieving before. They have made good progress but they still have some catching up to do.
		- or –	0	+ or ++
		Prior Value-added		

Research method

Methods of producing value-added estimations

To produce value-added estimations, assessors in education may choose between the multi-level modeling method and the ordinary least squares method. The former, also known as hierarchical linear modeling (e.g. Doran, 2003; Raudenbush & Bryk, 2002), is the most statistically valid method and it is preferable for research purposes. However, its procedure is technically complicated and its results may not be suitable for feedback purposes (Sharp, 2006). They might be too difficult for many users of assessment results, including teachers, school administrators, reading support specialists and many other educational practitioners that do not have a strong background in statistics. Therefore, there is a need to simplify the complicated assessment method or procedure to make value-added information easier or relatively friendly for them to understand.

Ordinary least square is another statistical method that school administrators, language teaching advisors, educational policy makers and other relevant educational professionals should not neglect because it is simpler and relatively friendly to users of assessment results. Its results are easier to interpret and to explain than the results of multi-level modeling. The method is not as powerful as multi-level modeling, as it is less flexible when handling data that has a hierarchical structure. However, in practice, there are also statistical findings suggesting that it has obtained the same results as the multi-level modeling method (Fitz-Gibbon, 1997). Therefore, as an alternative to the sophisticated statistical method of producing value-added information, the ordinary least square is a relatively friendly or comfortable statistical method to be used to produce school-based value-added in reading estimations in this paper.

Samples and measures

The data is comprised of information about 60 schoolchildren obtained from a primary school in Northern England. All of them have taken the PIPS assessment in Key Stage 2 and 58 of them participated in PIPS assessment when they were in Key Stage 1. The sample selection is in line with the intention of producing a school-based data analysis. Instead of using multi-school data, it is believed that data concerning schoolchildren of a single school is more likely to be available or familiar to school administrators and practitioners in education. The use of measures concerning schoolchildren of a single school for statistical analysis would facilitate understandings and interpretations of results produced in the analysis. Therefore, unlike most value-added analyses which are conducted with the use of data concerning students in different schools, the attempt

to make value-added analysis easier for readers to understand is a feature of this paper.

Samples collected from this school are selected mainly because the data set is conveniently available to the authors and they have been found to be suitable for producing the two types of value-added estimations. Its suitability is determined by the selection criteria that there is a positive and linear relationship between the two variables in each of the two linear regression models in this paper. In the selection process, the results of initial analysis of the equation and the parameters in each of the two statistical models have confirmed the existence of the relationships. To keep the presentation of statistical results in this paper simple, further details of the initial results will not be presented here.

The data contains 3 sets of measures collected by PIPS in England, namely RA, DA, and PA measures. These measures were not locally available when the data was collected. The first one is a set of standardized measures showing the results of individual schoolchildren in a reading test. The second one is a set of standardized measures showing the results of individual schoolchildren in a test of their developed abilities. The test contains curriculum-free assessment items of picture vocabulary and problems of positions. Hence, the DA measure is comprised of information concerning ability to deal with verbal and non-verbal tasks. The third one, PA, is a set of composite measures. The measure is the average of the standardized scores individual schoolchildren obtained in a reading test and in a math test completed in the previous assessment cohort, which was collected two years before assessment data of the current year was collected. To keep the measures on the same scale, each of them is converted as a set of standardized scores with the mean of 50 and the standard deviation of 10. Each of the standardization processes is completed with the use of results of all the schoolchildren who have completed the test(s) in the assessment cohort. The size of the estimated total population in each of the tests is over 40,000 schoolchildren. It means that each of the standardization processes was taken place with a large sample size. As data are processed to fit normal distribution, they can be used for comparisons with measures on different tests. Examples of different types of tests described in this paper and further information about the nature, type, meaning, and the quality of PIPS data and test administration are available from the website of the data provider (see PIPS, 2009 for detail).

Research questions

The research questions to address in this paper are:

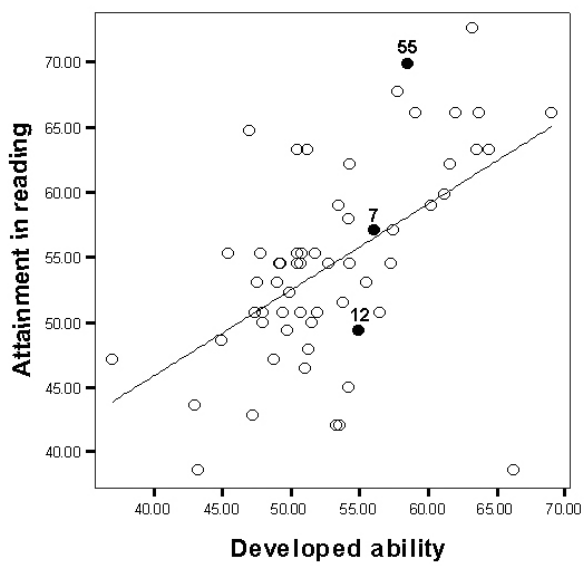
1. With reference to the PA or DA, what level of RA can one reasonably expect from individual students in the study?
2. What is the value-added in reading estimation?
3. Are the descriptions presented in the value-added grid useful for making interpretations of the two types of value-added in reading estimations produced in this paper?
4. Are there descriptions that have not been examined in this study?

To address the research questions mentioned above, the data analysis of this paper is implemented in two linear regression models. The procedures and results are reported in the text below. It is then followed by the examination and discussion about the use of the value-added grid as a tool for making interpretations of the value-added assessment results.

Data analysis of the first regression model

In the first regression model, data of the C_RA measure and the data of DA measure are used as paired assessment data. Figure 2 helps to explain the statistical estimation of E_RA and the calculation of value-added in reading estimation in the first regression model. The dots in the figure represent the current positions of individual schoolchildren in relation to their paired assessment data. As an example, dot number 7 represents the position of a schoolchild who has scored 56.20 in the DA measure and 57.07 in C_RA measure. Dot number 12 represents the position of another schoolchild who has scored 54.98 in the DA measure and 49.29 in C_RA measure. Similarly, the positions of each of the schoolchildren are plotted in the figure. Given that there is a positive and linear relationship between the variables in this model, it is appropriate to apply the ordinary least square statistical method to produce a best-fit linear regression line, as an outcome of the statistical model in the middle of the figure.

Figure 2: The relationships between attainment in reading and developed ability



Results of the first regression model

With the help of relevant computer software applications, it would not be difficult to draw a linear regression line that shows the general pattern of distribution of these dots. The straight line shows that schoolchildren with higher ability tend to have higher attainments in reading, while schoolchildren with lower ability tend to have lower attainments in reading. The line shows the attainments in reading that one can reasonably expect from the schoolchildren when the DA measure is taken into account. In this statistical model, the E_RA is statistically estimated when the DA measure of individual schoolchildren is taken into account. In practice, the value of E_RA of individual schoolchildren is an outcome of statistical estimation obtained from the operation of the equation as follows:

E_RA = Intercept + Gradient x DA

Then the value-added in reading estimation is calculated by taking away the E_RA from the C_RA of that schoolchild. Mathematically speaking, the value-added in reading estimation is calculated as a residual of the first regression model. Its procedure can be shown in the following equation:

CVA (of the 1st model) = C_RA – E_RA (i.e. when DA measure is taken into account)

Since the performance of individual students is compared with those of similar students in the assessment cohort, each of the value-added estimations reflects the relative advantage that an individual student has gained from learning and instruction. Such a gain is not attributed to personal development factors because it is likely that the

pace of development among students with a similar background (e.g. similar ability) tends to be equal when the sample size of the assessment cohort is large.

In this statistical model, the value-added in reading estimation is known as CVA in reading, which is a measure of the relative position of the schoolchild when compared with other schoolchildren of similar DA. A school-based example of this is presented in Figure 2. The E_RA of the schoolchild number 7, who has scored 56.20 in the DA measure, is 56.60. As the C_RA of the schoolchild is 57.07, the CVA in reading of the schoolchild is 0.47, which is close to 0. The result means that the schoolchild is doing as well as expected, when compared with other schoolchildren of similar ability.

With reference to Figure 2, the positive value-added estimations are represented by dots that are located above the regression line and the negative value-added estimations are represented by dots that are located below the regression line. An example is, the E_RA of schoolchild number 55, who has scored 58.53 in the DA measure, is 58.14. As the C_RA of the schoolchild is 69.90, the CVA in reading of the schoolchild is 11.76. The result means that the schoolchild is doing better than expected, when compared with other schoolchildren of similar DA. In contrast, the E_RA of schoolchild number 12 is 55.79. As the C_RA of the schoolchild is 49.29, the CVA in reading of the schoolchild is -6.50. The result means that the schoolchild is not doing as well as expected, when compared with other schoolchildren of similar DA.

Data analysis and results of the second regression model

In the second regression model, the data of the C_RA measure and the data of PA measure are used as paired assessment data. As there is a positive and linear relationship between the variables in the model, it is appropriate to apply the ordinary least square statistical method to produce a best-fit linear regression line. In this statistical model, the E_RA is statistically estimated, when the PA measure of individual schoolchildren is taken into account. In practice, the value of E_RA of individual schoolchildren is an outcome of statistical estimation obtained from the operation of the equation as follows:

$$E\_RA = Intercept + Gradient \times PA$$

The value-added in reading estimation of individual schoolchildren is calculated by taking away the E_RA from the C_RA of the respective schoolchildren. Mathematically speaking, its procedure can be shown in the following equation:

PVA (of the 2nd model) = C_RA – E_RA (i.e. when PA is taken into account)

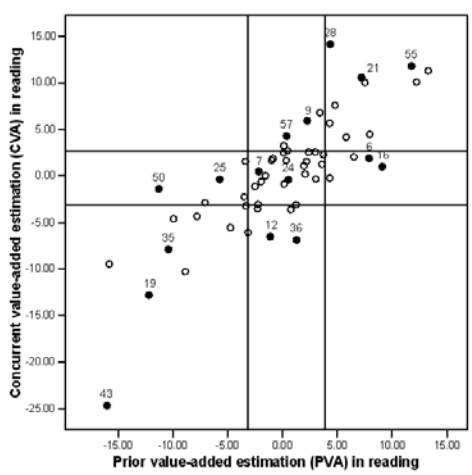
Since the relative performance of individual students in the previous assessment cohort is compared with the relative performance of their own in the current assessment cohort, each of the value-added estimations reflects the relative advantage that an individual student has gained from learning and instruction. In this statistical model, the value-added in reading estimation is known as PVA in reading, which is a measure of the progress of individual schoolchildren when compared with other schoolchildren of similar PA.

The use of a value-added grid as an assessment feedback tool

Figure 3 presents an illustration of the integration

of results of the two value-added estimations produced in this paper. For example, as in the text about the first regression model reported above, the CVA in reading of the schoolchildren numbers 7, 12 and 55 are 0.47, -6.50 and 11.76 respectively. Their PVA in reading, as produced in the second regression model but not reported in the text above, are -2.19, -1.14 and 11.77 respectively. When combining the results of the two value-added estimations, it is possible to present the position of each of them as a dot in Figure 3. The process continues until the positions of all the schoolchildren are presented in the figure. Then, as a replicate of the practice of Tymms & Albone (2002), horizontal and vertical lines are drawn to represent the 1st and the 3rd quartiles of the CVA and PVA measures. They divide the figure into 9 squares. To examine the validity of descriptions in the value-added grid, the results concerning the characteristics of schoolchildren inside the squares will be interpreted with reference to relevant descriptions in the grid.

Figure 3: Integration of the two types of value-added information



In the past, the attainments in reading of schoolchildren numbers 7 and 24 were probably as well as expected when compared with similar children in the school. Their reading progress has been a consistent trait over time. Therefore, their current attainments in reading are still as well as expected when compared with similar children in the school. To schoolchildren numbers 9 and 57, reading progress has also been a consistent trait over time. In the past, their attainments in reading were probably better than expected when compared with similar children in the school. As their learning progress has been steady over time, now their current attainments in reading are still better than expected when compared with similar children in the school. Schoolchildren numbers 12 and 36 are children who have retained a consistent trait in reading progress over time. In the past, their attainments in reading were probably not as well as expected when compared with similar children in the school. As their learning progress has been steady over time, currently their underachievement in reading becomes a continuing experience.

Schoolchildren numbers 21 and 55 have made excellent reading progress over time. In the past, their attainments in reading were probably as well as expected when compared with similar children in the school. As the extent of their reading progress has been large, their current attainments in reading are better than expected when compared with similar children in the school. Schoolchildren numbers 16 and 6 are also children who have made excellent reading progress over time. In the past, the attainments in reading of these children were probably not as well as expected when compared with similar children in the school. As the extent of their reading progress has been large, their current attainments in reading are as well as expected when compared with children in the school.

Schoolchildren numbers 19, 35, 25 and 50, have made little reading progress over time. In the past,

the attainments in reading of schoolchildren numbers 19 and 35 were probably as well as expected when compared with similar children in the school. As the extent of their reading progress has been small, the children are currently underachieving in reading. In contrast, the attainments in reading of schoolchildren numbers 25 and 50 were probably better than expected when compared with similar children in the school in the past. As the extent of their reading progress has been small, their current attainments in reading are only as well as expected when compared with similar children in the school.

Discussions, implications and limitations

So far, relevant descriptions in the value-added grid are found to be valid for making interpretations of the results. The descriptions are accurate and they have valuable implications for providing assessment feedback, for making instructional decisions and for supporting individual schoolchildren to learn. For example, schoolchildren in the squares in the middle column of the figure have made consistent progress. Reading support specialists, school administrators, language teachers and other educational practitioners might be satisfied to keep the consistent progress made by schoolchildren in the top middle square because they are gaining positions that are better than expected when compared with those of similar background. In contrast, these educational administrators, specialists and practitioners might have concerns about the little progress made by schoolchildren in the bottom middle square. Their attainments in reading are consistently not as well as expected over a period of time. They could be the group of schoolchildren who have specific learning difficulties (dyslexia) and require reading and language-related support arrangements.

Upon reflection, in the examination of relative advantage gained from learning and instruction, administrators and users of value-added assessment results should note that "even if prior attainment measures are available, educational practitioners may still want to look at concurrent general aptitude measures because there may be some pupils who have 'underachieved' throughout their school careers (Tymms and Henderson, 1995)". The statement provides further justification for the use of CVA simultaneously with PVA. The contrast between schoolchildren in these squares has implications for our understandings of variation in reading gains, which reveals the significance for users of value-added assessment to look for further information beyond the progress made.

Schoolchildren in the middle left-hand square were falling behind as a result of little progress made over time. As they were moved from positions where they were ahead of their peers, there is a need to find out the reasons that attribute to the little progress made and respond to the problem with relevant follow up actions. In contrast, acknowledgement and encouragement have to be given to schoolchildren in the middle right-hand square because of their success in catching up with their peers as a result of the excellent progress made over time.

Schoolchildren close to the bottom left-hand square have fallen behind and are now under-achieving, whereas schoolchildren close to the top right-hand square were moving further ahead of their peers. However, attention has to be given when making speculations about the positions of their prior attainments in reading, and such attention is essential for individual schoolchildren with extremely high or extremely low CVAs. Users of value-added assessment results need to be alert to such a limitation to the validity of the descriptions in the value-added grid. In practice, as it is clear that the CVA in reading of schoolchild number 43 is extremely low, one may question whether the child's attainment

in reading "was probably as well as expected" in the past. Instead, as the schoolchild's CVA is prominently smaller than his/her PVA, it would be better to speculate that the schoolchild "was probably under-achieving" in the past. Similarly, as it is clear that the CVA in reading of schoolchild number 28 is extremely high, one may also question whether the child's attainment in reading "was probably on track before". Instead, as the schoolchild's CVA is prominently greater than his/her PVA, it would be better to speculate that the schoolchild "was probably better than expected" in the past. The alternative interpretation may lead to a greater attention being given to find out the reasons for the change in relative positions, such as difficulties when settling in the school, an illness or family problems.

To sum up, the findings lead to the generalization that relevant descriptions of the value-added grid examined in this paper are generally accurate and useful for revealing the characteristics of the schoolchildren in this paper. Attention has to be given when using the descriptions to make speculations about the history of individual schoolchildren. In order to ensure the use of appropriate speculations that fit each of them, users of the grid might have to check and fine-tune the speculative descriptions on a case-by-case basis. It implies the importance for teachers, school administrators, reading support specialists and other educational professionals to equip themselves with a reasonably good understanding about the nature, meaning, importance and potential value of value-added estimations.

Readers have to be certain that the school-based value-added estimations reported in this paper should not be used to replace the national-project-based value-added estimations produced in PIPS, but they could be treated as supplementary or alternative value-added information that specifically reflects the relative learning gains of schoolchildren in the school. The information about the production of school-based value-added estimations and the illustration about the use of a value-added grid would serve as a relatively simple and friendly example to develop, update or enhance readers' understandings of value-added assessment and facilitate their interpretations about student learning gain.

The findings can make contributions to identifying which schoolchild is underachieving, which is on track, or is overachieving in relation to his peers in the school. However, the validity of descriptions in the value-added grid on itself does not ensure the appropriateness of its use. Given that it cannot change the quality of value-added assessment results, attention has to be paid to schoolchildren with extremely high or extremely low CVAs. Above all, readers should not ignore that the practical experience concerning the individual schoolchildren gained by school administrators or educational practitioners through observations or personal interactions have to be used for the checking and validation of findings produced in the statistical analysis.

The results of analysis have partially examined the validity of the descriptions in the grid as an assessment feedback tool and they have demonstrated the potentials of value-added information for improving learning and instruction. The general pattern of distribution of dots in Figure 3 is consistent with Tymms and Albone (2002), who suggested that the two types of value-added estimations correlated highly with one another. The descriptions in the top left-hand square and in the bottom right-hand square of the value-added grid were unexamined because of the lack of schoolchildren who fell in the two squares. Consequently, the results of school-based analysis in this paper cannot tell whether descriptions in the two squares of the grid are valid for the provision of feedback about the two school-based value-added estimations, or not. This is not surprising since PIPS

(1999) stated that the percentage of participants who belong to each of the two squares was less than 1%. The finding is providing support to the expectation that, as stated in text above, the usage of the descriptions in the two squares of the value-added grid is rather infrequent. It also implies that there are uncertainties to address the unexamined descriptions as a follow up investigation in the future through school-based value-added analysis with the use of the ordinary least square method. It might be worthwhile to consider alternative investigation strategies, such as identifying schoolchildren who belong to the two squares from the PIPS database and examine the appropriateness of the descriptions through in-depth interviews.

When reflecting on the process of evaluation, it is believed to be worthwhile to conduct an initial statistical analysis to evaluate the suitability of the data before the formulation of a linear regression model. The existence of a positive and linear relationship between the independent and dependent variables in the linear regression models is indeed a pre-requisite for the production of value-added estimations with the ordinary least square method in this paper. In fact, a high level of statistical significance is required to reveal the existence of such a relationship in the initial analysis. The reason for such a demanding sample selection criteria is to ensure the high quality of the data collection, given that the size of the sample is relatively small when compared with many other value-added assessment researches. Furthermore, there is also a need to address other technical issues in the initial analysis, such as the statistical significance of the parameter estimations, the proportion of explainable variance in the statistical model, the establishment of an assessment baseline, the implementation of a diagnostic check of distribution, standardization and normalization of data, and setting a confidence interval concerning the upper and lower limits of estimation. The efforts paid to these issues are important and useful for ensuring the high quality of the information to be used for the school-based analysis.

Conclusion

This paper provides an example of how to produce two types of value-added estimations with the ordinary least square method in a school-based evaluation of the relative advantage gained in reading by 60 schoolchildren in a primary school in England. The results are presented in a figure and they are interpreted by applying the descriptions in the value-added grid. The validity of the descriptions was examined. They were found to be applicable and useful for making interpretations about the characteristics of the schoolchildren on the basis of the value-added assessment results. However, attention has to be given when making speculations about the positions of individual schoolchildren's prior attainments in reading, such as those with extremely high or extremely low value-added estimations. Some descriptions in the grid were unexamined due to the lack of relevant samples that belong to the top left-hand and the bottom right-hand squares of the grid. It suggested an alternative investigation strategy for the follow up examination of the unexamined descriptions.

Results of the school-based analysis have shown that value-added assessment information has valuable implications for reflecting the academic learning gains, for making interpretations about the characteristics of individual schoolchildren, for making instructional decisions and for supporting individual schoolchildren to learn. The findings can make contributions for identifying which schoolchild is underachieving, which is on track, or is overachieving in relation to his peers in the school.

Acknowledgements

This study was kindly supported by projects funded by the Teacher Training Agency (TTA) and PIPS. The authors would like to give special thanks to the project leaders, David Moseley and Peter Tymms.

References

Curriculum Development Council [CDC] (2001). Learning to Learn: The Way Forward in Curriculum Development. Hong Kong: Curriculum Development Council.

Doran, H.C. (2003). Value-added analysis: A review of related issues. Paper presented at the Annual Meeting of the American Educational Research Association, ERIC document: ERIC accession number ED476134.

Doran, H.C. and Fleischman, S. (2005). Challenges of value-added assessment. *Educational Leadership, Vol. 63, No. 3*, pp.85-87.

U.S. Department of Education [ED] (2009). Secretary Spellings Approves Additional Growth Model Pilots for 2008-2009 School Year. Press Release, 8 January 2009, U.S. Department of Education. Retrieved on 1 May 2009, from http://www.ed.gov/news/pressreleases/2009/01/01082009a.html.

Department of Children, Schools and Families [DCSF] (2007). The Key Stage 1 to Key Stage 2 Value-added Measure. Retrieved on 29 November 2007, from http://www.dcsf.gov.uk/performancetables/primary_04/p3.shtml.

Education Bureau [EDB] (2007). School Value-added Information System. Hong Kong: Education Bureau (Hong Kong). Retrieved 1 December 2007, from http://svais.edb.gov.hk/default.asp.

Fitz-Gibbon, C. (1997). *Value Added National Project final report: Feasibility studies for a national system of value-added indicators*. London: The School Curriculum Assessment Authority, SCAA.

Kelly, S. and Monczunski, L. (2007).Overcoming the Volatility in School-Level Gain Scores: A New Approach to Identifying Value Added With Cross-Sectional Data. *Educational Researcher, Vol. 36, No. 5*, pp. 279–287.

Lee, K. and Landauer-Menchik, B. (2002). *The role of value-added assessment in Michigan's accountability system*. Policy Report No. 6, The Education Policy Center at Michigan State University.

Olson, L. (2004). Tennessee Reconsider Value-Added Assessment System. *Education Week*, *23*, pp.9-12.

Performance Indicators in Primary Schools [PIPS] (2009). Performance Indicators in Primary Schools (PIPS) website. Retrieved on 15 May, 2009, from http://www.pipsproject.org.

Performance Indicators in Primary Schools [PIPS] (1999). *Performance Indicators in Primary Schools Y2-Y8 newsletter: Issue number 8b*. Durham, England: CEM Centre, School of Education, University of Durham.

Raudenbush, S.W., & Bryk, A.S. (2002). Hierarchical Linear Models: *Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Tse, H. (2004). *An investigation into 2 approaches to predict pupil academic performance and calculate educational value-added*. Paper presented at the International Association for Education Assessment (IAEA) Conference 2004, from 13-18 June 2004 at Philadelphia, USA.

Tymms, P. (1999). Baseline assessment, value-added and the prediction of reading. Journal of *Research in Reading, Vol. 22, Issue 1*, pp.27-38.

Tymms, P. (1997). *The Value Added National Project Technical Report: Primary 4*. London: School Curriculum and Assessment Authority.

Tymms, P. and Albone, S. (2002). Performance Indicators in Primary Schools, In A. Visscher and R. Coe (eds.) *School improvement through performance feedback*. Lisse, Netherlands: Swets and Zeitlinger.

Tymms, P. and Fitz-Gibbon, C. (1995). Students at the front: Using performance indicators for professional development. *Educational Research*, *Vol. 37, Issue 2,* pp.107-122.

Tymms, P. and Henderson, B. (1995). *The Value Added National Project Technical Report: Primary.* London: School Curriculum and Assessment Authority.

Sanders, W.L. and Horn, S.P. (1998). Research findings from the Tennessee Value-added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, Vol. 12, No. 3,* pp.247-256.

Saunders, L. (1999). *Value-added measurement of school effectiveness: A critical review* Berkshire: National Foundation of Educational Research.

Sharp, S. (2006). Assessing value-added in the first year of schooling: Some results and methodological considerations. *School Effectiveness and School Improvement, Vol. 17,* No. 3, pp.329-346.

Author:

Dr. Harrison Tse, Assistant Professor,

Dept. of Curriculum & Instruction,
Hong Kong Institute of Education
E-mail: htse@ied.edu.hk

Mr. Stephen Albone
The Curriculum, Evaluation and Management Centre,
Durham University
E-mail: S.A.Albone@cem.dur.ac.uk

Received: 16.9.08, accepted 21.10.08, revised 24.10.08, further revised 27.5.09