Assessment of Creative Writing in Vietnamese Primary Education

Patrick Griffin
The University of Melbourne
Australia

Phan Nguyet Anh
The University of Melbourne
Australia

The teaching and assessment of essay writing at primary schools throughout Vietnam is regulated by the Ministry of Education and Training. The analytical error-recognition method of assessment, however, does not facilitate direct interpretation of students' writing competence. In this study, which involved samples of Grade 5 students in five provinces in Vietnam, a combination of traditional and partial credit scoring rubrics was developed to enable data analysis using the Rasch model. Based on such analysis, a continuum of writing ability at Grade 5 level was identified and a mastery level defined in terms of writing skills. The study has implications for possible changes in future assessment and marking schemes.

Key Words: Assessment, Measurement techniques, Criterion referenced tests, Creative Writing

The purpose of this study was to examine the creative writing ability of Grade 5 students in Vietnam, and to identify and interpret levels of developmental progress in writing ability. The study also examined the scoring practices of teachers assessing writing of year 5 pupils in Vietnam and then linked these practices to the identification of developmental levels and teaching intervention practices.

The study arose from a larger investigation of literacy and numeracy in five provinces in Vietnam (Griffin, 1998) which examined the quality of education being delivered to primary students. Of concern was the process of identifying factors that contribute most strongly to quality education.

From this arose the objective of identifying how educators and planners in Vietnam might best manage these factors to promote success in schools, and how curriculum intervention might best occur at classroom, school and system level to improve basic skill performance of primary students at the point of transition to secondary school. Many studies have considered the relative importance of different "inputs" to some measure of school-based student performance, usually based on a specific subject test (such as language and/or mathematics) or on successful attainment of a benchmark academic level (such as a school-leaver certificate or mastery of specific content) (Greaney, V., Khandker, S.R., & Alam, M. 1990; Ahlawat, K. and V. Billeh, 1994; Ahmed, S. M. and L. T. Salih, 1996; Al Nhar, T. 2000; Al-Nhar, T. and V. Billeh, 1997; Castro, M. H., 2000; Chinapah, V.,2000; Chinapah, V., 1999; Falayajo, W., G. Makoju, 1997; Khaniya, K.,1999; Machingaidze, T., P. Pfukani 1998 ; Narros, S. and K. A. Mohammed, 1998; Odeh, D. and M. H. Kizilbash, 1998; Voigts, F. 1999). Inputs of common interest in such studies include school infrastructure, teacher training, teacher supervision and incentives, curriculum, students' physical well-being, textbooks and other pedagogical materials, and family and community context among others. This paper addresses the assessment of student creative writing at or near

Patrick Griffin, the Chair of Education (Assessment) at the University of Melbourne, Director of the Assessment Research Centre, Head of Department of Learning and Educational Development, a project team leader for UNESCO in southern Africa and a World Bank consultant in Vietnam. **Anh Nguyet**, a professional translator and interpreter who completed her postgraduate program at the University of Melbourne in the field of assessment.
Correspondence concerning this article should be addressed to Department of Learning and Educational Development, University of Melbourne, Victoria 3010, Australia. Electronic mail may be sent to p.griffin@unimelb.edu.au

the end of Grade 5 or at the end of primary school education, given the control placed on the instruction and assessment of student writing in Vietnam.

Vietnam

The Socialist Republic of Vietnam has an area of about 330,000 square kilometres, and is bounded by the Gulf of Tonkin and the South China Sea, China to the north, Cambodia and Laos to the west, between the 9th and 26th parallels north. The capital, Hanoi, is in the main northern region. The ancient capital, Hue, is in the main central provinces region, and Ho Chi Minh City (formerly Saigon) is in the main southern region. The total population exceeds 70 million with about 85% ethnic (Kinh) Vietnamese, and the remainder from many ethnic minority groups. The official language is Vietnamese, but French, English, Khmer and tribal languages are spoken.

After the reunification of Vietnam in 1975-76, the government sought to develop a single, uniform and unified education system from the separate systems that had operated in the previous 30 years. There was a sustained and major program to increase the proportion of literate adults, to increase the proportion of children receiving the initial five years of basic education, and to expand access to tertiary studies.

The regions chosen in this study.

The present project focused on five provinces, targeted by the World Bank for special investigation and development. The study involved a grade-based population in five provinces: Ha Noi, Yen Bai, Thanh Hoa, Quang Nam and Vinh Long. These provinces represent a northern urban industrialised province (Ha Noi); a central urban and rural province (Quang Nam); an isolated mountain province (Yen Bai); a relatively poor, rural farming province (Thanh Hoa); and a southern, Delta province where the mainstay of the economy is fishing and rice (Vinh Long). Generalisations, as much as they can be made, can be offered only at provincial level rather than at a national level.

School education in Vietnam

Under current arrangements, schooling in Vietnam commences with pre-primary education (where places are available) provided by crèches (birth to three years) and kindergartens (3 to 6 years). This is followed by five years of

primary schooling (nominally free and compulsory). The school year runs from September to June and the medium of instruction is Vietnamese. At the end of primary education, a certificate of Primary Education (Bang Tieu Hoc) is issued. Graduates then progress either to lower secondary school, (Basic General Education Level 2) for four years or to Vocational Training school (for three years). Graduates for this new level receive either the Certificate of Lower Secondary Education from the former or a Professional Certificate from the latter. Students can then proceed to one of four strands of education. These are upper secondary education (Diploma of General Education), Secondary Technical Education (Diploma of Secondary Technical Education), Secondary Vocational School (Diploma of Secondary Vocational Education), or Vocational Training School (Professional Diploma). On completion of 12 years of education, entry to the next stage is via examination set by the Ministry of Education and Training (MoET). These include University (Phase 1) or Junior College.

Primary Education in Vietnam

In primary education, literacy, which for the purposes of this study is referred to as "Vietnamese language", is considered the most important subject. It accounts for almost 50% of classes per week for students at Grade 1, over 40% of the number of classes at Grades 2 and 3, and 30% of the time designated for the whole curriculum at Grade 4 and Grade 5 (Ministry of Education, 2000). The general distribution of time for subjects in the curriculum is shown in Table 1.

Textbooks in Vietnamese language, as well as in other subjects, are nationally prescribed by MoET. The curriculum guidelines, also issued by MoET, provide detailed instruction on the amount of time to be spent by teachers and students on each section of a 40 minute class (Ministry of Education, 2001). Teachers are encouraged to improve these lesson plans, but permission is needed through the teacher assessment program to change or replace them. Students are taught pronunciation, parts of speech, and reading and writing skills using short proverbs, poems or Vietnamese stories. After each text, there are questions and exercises that are expected to consolidate what is learned from the text.

Teaching writing at primary education level in Vietnam.

Writing is taught in conjunction with reading, grammar, vocabulary and spelling using set textbooks for each grade at the primary education level. The textbooks in Vietnamese

Table 1. *Classes per week taught by grade and subject at primary education level*

Subject	Grade Level				
	1	2	3	4	5
Vietnamese language	11	10	9	8	8
Mathematics	4	5	5	5	5
Ethics	1	1	1	1	1
Natural and Social Studies	1	1	2		
- Science				2	2
- History				1	1
- Geography				1	1
Technology	1	2	2	2	2
Music	1	1	1	1	1
Fine Arts	1	1	1	1	1
Physical Training	1	2	2	2	2
Health Education	1	1	1	1	1
Total:	22	24	24	25	25

language are structured so that instruction is integrated and students only do writing tasks on a certain topic after they have read about the topic, and learned relevant spelling, grammar points and necessary vocabulary for that topic.

Children learn how to write individual letters and words at Grade 1. At Grade 2, writing at sentence and paragraph levels is taught. Each topic is covered in one class, mostly in the form of asking students to answer questions based on a picture. Grade 3 teachers spend two classes on each topic. The first class focuses on oral recount of stories that the students have learned, and the second class is for students to re-write the same stories in their own words. Creativity is encouraged at this level in the form of use of synonyms and paraphrasing.

At Grades 4 and 5, students start to write their own essays in the form of a narration or description. Three classes are spent on each topic, with the third class used for comments on writing tasks done by students. Students in Grade 4 are given lessons in basic essay structure, consisting of an introduction, body and conclusion. Cohesion at the essay level is obtained through practising the use of connecting words and ideas and applying chronological or other logical sequences. Common tasks include description of an object, an animal or a landscape. At Grade 5, the requirements of creative writing extend to development of lively, expressive essays. The subjects to be described are generally restricted to people and activities in everyday contexts. Students are given practice in extending simple sentences into compound sentences, and in applying such figures of speech as comparison and association. They are also introduced to social letters and application letters. Before each writing task,

teachers help students construct a detailed outline. In the "comment class", teachers remark on essay structure, cohesion, and use of vocabulary and syntax, and help students to correct common errors in their writing.

Assessment of Vietnamese language.

Assessment is mandated and is carried out in accordance with MoET Circular 15/GD-DT dated 2 August 1995. There are two types of tests: Regular tests (R) and periodical tests (P). Each contributes to an overall assessment mark for the subject, but in different ways. There are at least four regular (R_i) tests (i=Sep, …, May) for each semester (k=1, 2). There is one test for reading, one for dictation, one for vocabulary and syntax, and one for essay writing. These tests can be in the form of 15-minute written tests, oral tests, or practice exercises.

At Grade 1, there are two periodic tests in the second semester of the school year. Students from Grade 2 to Grade 5 are required to do two periodical tests each school semester (P_m where m=1, 2). Each periodic test consists of a reading component and a writing component. The writing component at Grades 2 and 3 includes a dictation exercise and a 25-minute composition. At Grades 4 and 5, students do a dictation exercise, one or two exercises in vocabulary and syntax, and a 40-minute essay. The result of each periodic test is calculated by averaging the marks for the reading and writing components. Any decimal is rounded up to the nearest whole mark if the writing component is marked higher than the reading component, or rounded down if the reading component is marked higher.

Achievement as at mid-semester 1, (S_{1m}), is based on September and October assessments where n_R is the number of regular assessments for the first semester:

$$S_{1m} = \frac{1}{(n_R+2)}\left(\sum_{i=Sep}^{Oct} R_i + 2P_1\right)$$

Achievement at the end of semester 1 (S_{1e}) combines all regular assessments from November to December and adds the end of semester assessment (January):

$$S_{1e} = \frac{1}{(n_R+2)}\left(\sum_{i=Nov}^{Dec} R_i + 0.5R_{Jan} + 2P_2\right)$$

- Overall assessment for Semester 1 combines the mid and end of semester assessments using a weighting process:

$$S_{1\cdot} = \frac{1}{3}\left(S_{1m} + 2S_{1e}\right)$$

Achievement for mid-semester 2 (S_{2m}) is based on January and February assessments following similar procedures as for semester 1:

$$S_{2e} = \frac{1}{(n_R+2)}\left(\sum_{i=Jan}^{Feb} R_i + 2P_3\right)$$

- Achievement as at the end of Semester 2:

$$S_{2e} = \frac{1}{(n_R+2)}\left(\sum_{i=Mar}^{Apr} R_i + 0.5R_{May} + 2P_4\right)$$

- Achievement for the whole of Semester 2:

$$S_{2\cdot} = \frac{1}{3}\left(S_{2m} + 2S_{2e}\right)$$

Achievement for the whole school year:

$$S_{\cdot\cdot} = \frac{1}{3}\left(S_{1\cdot} + 2S_{2\cdot}\right)$$

The marking method is analytical, with specific scores assigned for each sub-component. In the teaching and assessment of reading, students are marked for their reading speed and their answers to comprehension questions. Students are marked on a scale ranging from one to ten, with one being the lowest mark and ten being the highest. No decimal point is used in the mark. Five marks are available for the essay, three marks for the dictation and two marks for the exercises. One mark is deducted for every three spelling mistakes in the dictation, although identical spelling mistakes are counted only once. For the essay, the introduction is given one mark, the conclusion one mark, and the body is marked out of three.

A student's overall achievement in the subject is calculated four times in a school year (at the middle and end of each semester, giving increasing weight to the assessment as the year progresses) according to the following formulae:

This conflates to a single annual scoring formula that weights performances at the end of the year more heavily than performances at the start of the year, presumably based on the assumption that assessments at the end of the year were more demanding than assessments at the beginning. While this is most likely true, the procedure adopted in this article weights the items by their capacity to discriminate between students, not the difficulty of the task or the time of the year. The items are in fact not 'weighted' in the sense of the normal use of the term at all. While the Rasch model automatically considers the relative difficulty, it is the discrimination that affects the relative weighting or contribution to the estimates of the pupil ability (Wu and Adams , 2005).

this paper weights automatically for the difficulty or the amount of demand that each sub-task places on the student and estimates the ability of the student from the tasks that they perform and the quality shown in each performance:

$$S_{\cdot\cdot} = \frac{1}{9(n_R+2)}\left\{\sum_{i=Sep}^{Oct} R_i + 2\sum_{i=Nov}^{Feb} R_i + 4\sum_{i=Mar}^{Apr} R_i + R_{Jan} + 2R_{May} + 2P_1 + 4(P_2+P_3) + 8P_4\right\}$$

Each score (R_i and P_m) range is limited to values between zero and ten and the final score is converted to a grade as shown in Table 2.

Table 2. *Score Conversion to Student Classifications*

Classification	Grade	Score Range
Excellent	A	9.0 - 10
Good	B	7.0 - 8.9
Average	C	5.0-6.9
Weak	D	4.9 or below

Method

Participants

In each of the five Vietnamese provinces, a sample of between 11 and 15 schools was randomly selected and, within each of the sampled schools, a single class of students was also randomly selected as required by the supervising government department. This yielded a total of 2032 students in 67 schools from the five provinces. The structure of the sample is shown in Table 3. The unit of analysis was the student despite the fact that a single intact class was selected from each school.

Table 3. *The Sample of Schools and Pupils*

Province	Schools	Pupils
Ha Noi	15	504
Yen Bai	14	442
Thanh Hoa	14	286
Quang Nam	11	386
Vinh Long	13	414
Total		2032

Assessment and Materials

Consistent with the Vietnamese curriculum for Grade 5, the elements of writing used for the assessment in the present study included story structure, use of language, spelling and syntax, and creative writing. All parts of the assessment system were allocated a scoring rubric that indicated different qualities of performance between students within that strand. Items were scaled according to the amount of ability needed to demonstrate the performance described for each score level.

Using a Rasch analysis (Rasch, 1960; Wright & Masters, 1982), the score assigned to an element or a component of the task is a means of coding only relative levels of performance quality and is not interpreted directly. Rasch modelling supports the interpretation of the student performance within a criterion-referenced framework. The same rule applies to dichotomously scored tasks and to rating scale items.

Typical of the requirements of Grade 5, the students were asked to write a description of a close friend or a member of the family:

There are a lot of people who are close to you in your life (such as your father, mother, your grandfather, grandmother, brothers, sisters, teachers or friends and so forth). Write an essay about one of these persons.

The project began with training sessions in Hanoi for the field workers, who collected data and supervised the test administration, and for the essay markers who were responsible for marking the scripts.

The construct of creative writing

Part of the training program consisted of identifying the likely progression and development levels of writing ability for Grade 5 students. Describing the anticipated levels of performance, and setting these down in a developmental framework, is a method of hypothesising the underlying construct to be measured. Table 4 presents the hypothesised construct underpinning the design of the assessment of creative writing. Setting the hypothesised construct a priori means that both students and items (or writing tasks) are expected to have a location on this variable. This was a novel approach to both the teaching and assessment of writing, and a departure from official policy for the assessment of writing in Vietnam. A compromise approach emerged in the definition of the scoring rubrics.

Scoring the essay

As shown in Table 5, the essays were marked using rubrics linked to the structure, style and content aspects of the essay, and by counting the number of errors made by students in linguistic aspects, including vocabulary, spelling and syntax. These approaches were consistent with current practices used by teachers in arriving at the periodic, annual, and regular assessments of Grade 5 student writing. The error-counting method was also applied to the criteria of cohesion and creativity.

The error-based marking method makes it difficult to analyse and interpret the scores in any detailed diagnostic manner. This was clear in the case of creativity and cohesion where "errors" and "ideas" were difficult to define and even harder to count. Markers, however, applied the routine marking rules that were regularly used with Vietnamese writing assessment and this may have increased the reliability of the scoring.

Data Analysis

The scores for the writing exercise were treated as 13 separate test items scored using a partial credit approach, with the highest score for each item reflecting the highest quality performance for that item and the lowest score reflecting the lowest quality performance. Scoring each item in this manner treats them as 13 independent polychotomous items, in which each student, n, has a writing ability θ_n and each item has a set of difficulty parameters $\delta_{i1}, \delta_{i2}, \delta_{i3} \ldots \delta_{ik}$ representing the difficulty of attaining each of the scores from 1 to k for item i. Each of these parameters governs the likelihood of a student with ability, θ, obtaining a score of k rather than k-1. The analysis models the relationship between student writing ability and the difficulty parameters of each of the 13 items. The Rasch model estimates student writing ability independent of which particular items are used for the estimation. The natural logarithm of the odds of achieving a

Table 4. *Proposed construct of creative writing for year 5 students in Vietnam*

Level	Expected Performance
3	High level writing involves the use of creative expression, correct use of language and an ability to present the story in a bright and entertaining way.
2	Medium level writing involves correct use of script, words and structure of the story such that the beginning, middle and end are clearly presented in a correct fashion.
1	Low level writing involves a lack of structure to the story, poor spelling and sentence structure; ideas are unconnected. Little reference to the characters in the story or the plot.

Table 5. *Scoring the Essay*

Item No.	Item Name	Response Description	Variable Map Designation	Code
1.	Introduction	None		0
		Unclear, inserted in the body	1.1	1
		OK (short, in one sentence)	1.2	2
2.	Appearance	No ideas		0
		Several ideas	2.1	1
		Enough ideas to visualise the person	2.2	2
3.	Characteristics	No ideas		0
		Several ideas	3.1	1
		Truthful and persuasive details	3.2	2
4.	Tone of story	Inappropriate or inconsistent tone		0
		The plot of the story is told	4.1	1
		The story sounds lively and credible	4.2	2
5.	Structure	No clear division or not enough parts		0
		With 3 clear parts	5.1	1
6.	Style	Inappropriate		0
		Appropriate	6.1	1
7.	Conclusion	No conclusion		0
		Inadequate conclusion not separated from the body	7.1	1
		Reflecting truthful impressions and feelings about the person	7.2	2
8.	Handwriting	Incorrect style		0
		Correct style	8.1	1
		Correct size	8.2	2
		With enough strokes	8.3	3
		With even and connecting strokes	8.4	4
9.	Spelling	Over 5 errors		0
		4 - 5 errors	9.1	1
		Under 3 errors	9.2	2
10.	Word use	> 3 errors		0
		3 errors	10.1	1
		1 - 2 errors	10.2	2
		0 error	10.3	3
11.	Syntax	> 3 errors		0
		3 errors	11.1	1
		1 - 2 errors	11.2	2
		0 error	11.3	3
12.	Cohesion	> 4 errors		0
		4 errors	12.1	1
		3 errors	12.2	2
		2 errors	12.3	3
		1 error	12.4	4
		0 error	12.5	5
13.	Creativity	0 ideas presented		0
		1 idea presented	13.1	1
		2 or more ideas	13.2	2

specific score of k rather than k-1 is obtained from the simple relationship

$$\ln\frac{p_k}{p_{k-1}} = \ln\frac{n_k}{n_{k-1}}$$

where p_k and p_{k-1} are the proportions of students scoring k and k-1 respectively.

Most items were scaled using IRT (Item Response Theory) scaling methodology. With the One-Parameter (Rasch) model (Rasch 1960) for dichotomous items, the probability of selecting category 1 instead of 0 is modelled as

$$P_i(\theta) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)} \tag{1}$$

where $P_i(\theta)$ is the probability of person n to score 1 on item i. θ_n is the estimated latent trait of person n and δ_i the estimated location of item i on this dimension. For each item, item responses are modelled as a function of the latent trait θ_n.

In the case of items with more than two (k) scoring categories (as for example wit a maximum x=score greater than 1) this model can be generalised to the *Partial Credit Model* (Masters and Wright, 1997).[1] The Partial Credit Model developed by Masters (1982) is an extension of the Simple Logistic Model, and overcomes the restriction to dichotomous scoring. The model was developed by estimating parameters for the difficulties associated with a series of performance levels within each item. Masters (1982) argued that the difficulty of the kth level in an item governs the probability of responding in category k rather than in category k - 1. The probability of person n of completing the kth level is specified by Masters (1982; 158) as:

$$P(X_{ni} = x) = \frac{\exp\sum_{k=0}^{x}(\theta_n - \delta_{ik})}{\sum_{h=0}^{m_i}\exp\sum_{k=0}^{h}(\theta_n - \delta_{ik})} \tag{2}$$

The model estimates the probability of a person n scoring x on the mi performance level of item i as a function of the person ability on the variable being measured and the difficulties of the mi levels in item i. The observation x is a *count* of the successfully *completed* item levels, while only the difficulties of these completed levels appear in the numerator of the model. The model provides estimates of person ability θ_n and item step level difficulty δ_{ik} and

---

[1] An alternative is the Rating Scale Model (RSM) which has the same step parameters for all items in a scale (see Andersen, 1997).

$P_{xi}(\theta)$denotes the probability of person n to score x on item i. θ_n denotes the person's position on the latent trait, the item parameter δ_{ik} gives the location of the item step, k, on the latent continuum and δ_{ik+1} denotes an additional step parameter.

Item fit was assessed using the weighted mean-square statistic (infit), which is a residual based fit statistic. Weighted infit statistics were reviewed both for item and step parameters. The ACER Conquest software (Wu, Adams and Wilson, 1997) was used for the estimation of item parameters and the analysis of item fit.

Given that the 13 items (or sub-tasks) have variable maximum scores, the partial credit model (Wright & Masters, 1982) using the computer program Quest (Adams & Khoo, 1995) was used to derive the estimates of item difficulty and student writing ability. (1)

Results and Discussion

The Variable Map

Estimates of θ_n and of the difficulty parameter δ_i were simultaneously plotted on a chart called a variable map that illustrates the relative position of students against the difficulty levels of score points assigned to each of the 13 items. These are shown in Figure 1. Where the student ability was at the same level as the difficulty of scoring a specific number of points on an item, then the odds that the student would score at least that amount for the item were 50/50. The logarithm of these odds was zero, indicating that there was no difference between the ability of the student and the difficulty of scoring at this level.

Two indicators of accuracy were used. The first was the standard error of measurement for each of the item difficulty estimates. The second was an indicator of the extent to which the data fit the Rasch model. This statistic is the mean squared difference between the modelled difficulty and the observed difficulty of each score point, weighted by the variance of the assigned scores. This is called the INFIT mean square. The expected value of the INFIT is 1.0 and accepted range of these values lies between 0.77 and 1.3 (Adams & Khoo, 1995), and when the item set sits completely within these limits it is taken as evidence of a single, dominant dimension underpinning the performances of the students.

The underlying construct, hypothesised in Table 4, can be examined using the variable map. Items can be seen to group together at different points along the scale. Of most interest, however, was determining whether these clusters had something in common. This is a matter of interpretation and is
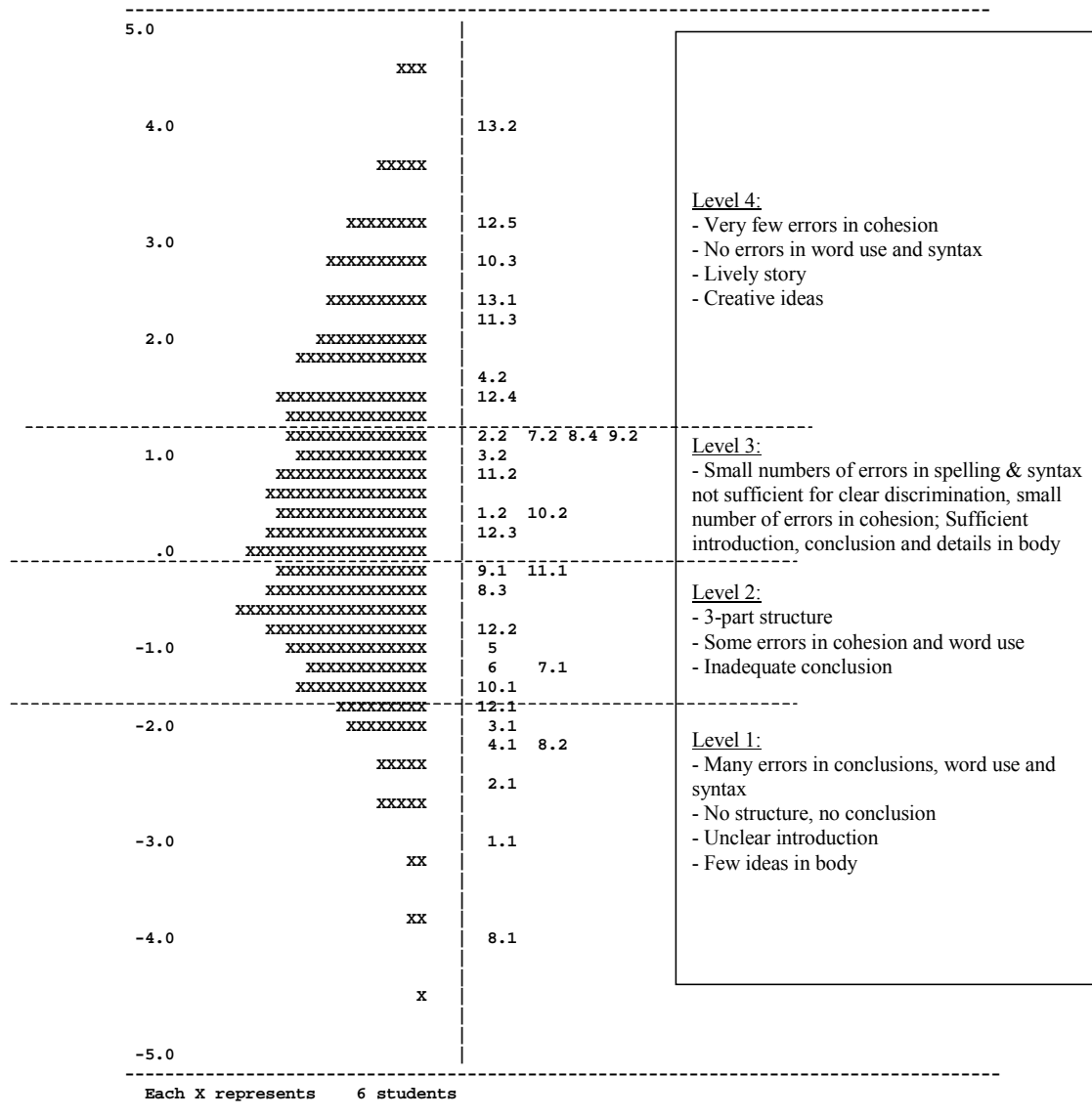
```
------------------------------------------------------------------------------
  5.0                           |
                                |
                    XXX         |
                                |
  4.0                           | 13.2
                                |
                  XXXXX         |
                                |
                                |                  ┌──────────────────────────
  3.0           XXXXXXX         | 12.5             │
                                |                  │  Level 4:
                XXXXXXXXXX       | 10.3             │  - Very few errors in cohesion
                                |                  │  - No errors in word use and syntax
                XXXXXXXXXX       | 13.1             │  - Lively story
                                | 11.3             │  - Creative ideas
  2.0          XXXXXXXXXXX       |                  │
               XXXXXXXXXXXX      |                  │
                                | 4.2              │
             XXXXXXXXXXXXXX      | 12.4             │
               XXXXXXXXXXX       |                  │
             XXXXXXXXXXXXX       | 2.2  7.2 8.4 9.2 └──────────────────────────
  1.0           XXXXXXXXXXXX     | 3.2                 Level 3:
                XXXXXXXXXXXX     | 11.2                - Small numbers of errors in spelling & syntax
               XXXXXXXXXXXXX     |                     not sufficient for clear discrimination, small
               XXXXXXXXXXXX      | 1.2  10.2           number of errors in cohesion; Sufficient
               XXXXXXXXXXXX      | 12.3                introduction, conclusion and details in body
   .0         XXXXXXXXXXXXXX     |
               XXXXXXXXXXXX      | 9.1  11.1
               XXXXXXXXXXXXX     | 8.3                 Level 2:
              XXXXXXXXXXXXXX     |                     - 3-part structure
               XXXXXXXXXXXX      | 12.2                - Some errors in cohesion and word use
 -1.0           XXXXXXXXXXXX     | 5                   - Inadequate conclusion
                XXXXXXXXXX       | 6    7.1
                XXXXXXXXXX       | 10.1
                 XXXXXXXX        | 12.1
 -2.0            XXXXXXX         | 3.1                 Level 1:
                                | 4.1  8.2            - Many errors in conclusions, word use and
                  XXXXX          |                     syntax
                                | 2.1                 - No structure, no conclusion
                  XXXXX          |                     - Unclear introduction
 -3.0                           | 1.1                 - Few ideas in body
                    XX           |
                                |
                    XX           |
 -4.0                           | 8.1
                                |
                     X          |
                                |
 -5.0                           |
------------------------------------------------------------------------------
       Each X represents    6 students
```

Figure 1. Variable map of writing analysis

closely aligned with the development of profiles outlined by Griffin (1990), but differs from other approaches to profile development or definition as exemplified by the reading levels for the PISA project (OECD Program for International Student Assessment, 2001). It requires a qualitative form of "artistic" interpretation following an audit of skills involved in attaining specific score points on assessment tasks.

Griffin (2001) illustrated how these charts can help to define developmental levels of students and how an instructional intervention can be posited. The variable map shows that students could be located at about the same level on the variable as a group of items. These students have about a 0.5 probability of success on those items, a higher probability on items below them on the variable, and a lower probability on items above them on the variable. Their location proximal to the set of items is a kind of "transition point". If a student were to improve a little, he or she would have a better chance of succeeding on items in this group. If a student were to regress, he or she would have a probability lower than 0.5 of succeeding on items in this group. Turner

(2005) illustrates the empirical approach to this. A content analysis of the skills required to succeed on the set of items within each of these groups also describes the kind of skills being demonstrated by students at this location on the variable and it helps to identify the kind of instruction needed to progress the student on the variable.

An important characteristic of this approach is the reconstruction of the proposed dimensions underpinning the assessment. If the skills audit "back translates" to match or approximate the originally proposed underpinning variable used to design and construct the test, it can be used as partial evidence of construct validity. When this is linked to the index of item separation, we have two pieces of evidence for construct validity (Wright & Masters, 1983). The technique has been used sparingly but has emerged in several international studies. Greaney and others used the procedure in their report on the Bangladesh testing, the "Education For All" project (Greaney, Khandker & Alam, 1995) in which they cited Griffin's and Forwood's (1990) application of this strategy in adult literacy.

Calibration.

Tables 6 and 7 present the calibration statistics for the assessment of student writing samples. Each of the parameters was estimated using the Rasch model, as reported in Table 6. The table presents each score step difficulty estimate, δ_{ij}, the standard error of measurement associated with each score step difficulty estimate (SE), and the extent to which the pattern of student responses associated with each score step fits the Rasch partial credit model (INFIT).

Some things are notable in Tables 6 and 7. The

measurement errors are very small due to the large sample size; the INFIT values are all within the range of 0.7 to 1.3; the variance of item difficulty levels is 1.56 with a reliability of item separation of 0.99; the mean item INFIT is 0.99 with a variance of 0.02; the mean student ability estimate is 0.32, indicating that the student ability level was slightly higher than the difficulty of the overall (easy) task. The variance of the student ability estimates is 2.12, almost equivalent to the variance of item difficulties, and this indicates that the spread of task difficulties was relatively well matched to the range of student abilities. The reliability of the student separation index is 0.89. The mean squared INFIT index is 1.05 with a variance of 0.29. This evidence, taken together, indicates that the test was well matched to the student population, that a single dominant dimension underpinned the task, and that the test was successful in separating the students on the basis of ability (i.e. that it possessed acceptable criterion validity) as well as providing support for claims of construct validity. With regard to construct validity, however, there was no external evidence of the nature of the criterion except that it was correlated to reading skills as measured by the other parts

Table 7. *Test Characteristics and Parameter Summary*

	Mean	Variance
Item separation	0.99	
Item INFIT	0.99	0.02
Item Difficulty	0.00	1.56
Student Ability	0.32	2.12
Student Separation	0.89	
Student INFIT	1.05	0.29

Table 6. *Item Parameter Estimates - Difficulty, Measurement Error and INFIT*

Item	δ_1	SE_1	δ_2	SE_2	δ_3	SE_3	δ_4	SE_4	δ_5	SE_5	INFIT
1	-3.06	0.19	0.48	0.08							1.08
2	-2.41	0.13	1.13	0.12							0.91
3	-1.84	0.13	1.04	0.13							0.86
4	-2.13	0.13	1.79	0.12							0.78
5	-0.92	0.06									0.85
6	-1.25	0.06									0.86
7	-1.09	0.09	1.11	0.11							0.92
8	-3.88	0.22	-2.16	0.14	-0.37	0.1	1.1	0.11			1.14
9	-0.03	0.11	1.16	0.09							1.28
10	-1.34	0.09	0.54	0.09	2.78	0.09					1.12
11	-0.09	0.09	0.85	0.1	2.33	0.11					1
12	-1.72	0.13	-0.79	0.09	0.21	0.08	1.52	0.1	3.11	0.13	1.17
13	2.47	0.16	4	0.22							0.99

of the test to the extent of 0.49.

Interpreting the variable map

The Rasch model was used to calibrate simultaneously dichotomous items and partial credit items for plotting on the variable map shown in Figure 1. On the left of the map is a scale ranging from -5.0 to +5.0. This is the logit scale of the Rasch model and represents the logarithm of the odds of a student scoring at least k rather than k-1 within each item. It also represents the ability estimates of students, represented by the Xs, and the difficulty of the item score level, represented by the x.y notation on the right of the map. In the x.y notation, the x represents the item number and y represents the score obtained, so 3.2 represents a score of 2 on item 3. The descriptions in the column to the right of the map indicate the interpretation of the clusters of score rules for each of the score points. A skills audit of the clusters of items (score steps) yielded the descriptions of the levels of competence in writing. Four main clusters are shown on Figure 1, and these were interpreted using the meaning of the score points (x.y as shown on the map and explained in Table 5).

Essays written by students in the bottom cluster had an unclear structure and a basic introduction (1.1) but no clear conclusion to the story. Although some ideas about the character and appearance of the person described were included (3.1 and 4.1), the vocabulary was poor. There were spelling mistakes and syntax errors in these essays. Cohesion was weak in the writing (12.1) and command of script was poor.

In the second cluster, basic structure was present in the essay (5.1) although the conclusion was inadequate (7.1). There were errors in language use with regards to both lexis (10.1) and syntax (11.1). Cohesion was better (12.2), but ideas were not clearly expressed.

Students in the third cluster showed a clear structure in essay writing with an appropriate introduction (1.2) and an adequate conclusion (7.2). Writing was logical and coherent (12.3). Sufficient details enabled visualisation of appearance (2.2) and persuasive description of character (3.2). Spelling and syntax were acceptable (9.2 and 11.2), and scripts met the required standards of sufficient and even strokes (8.4).

The fourth cluster consisted of students who could use the language correctly (10.3 and 11.3) and creatively. Their writing was coherent and logical (12.5) with an imaginative story line (4.2). These students were able to use visual imagery and creative expression (13.2) to tell a story in a lively and engaging style.

Table 8. *Comparisons of the Hypothesised and derived construct*

Original Construct	Derived Construct
3. High level writing involves the use of creative expression, correct use of language and an ability to present the story in a bright and entertaining way.	4. Extensive use of visual imagery and creative expression to tell a story, creative use of language and story structure. The writing is coherent and logical with an imaginative story line and a lively and engaging style.
2. Medium level writing involves correct use of script, words and structure of the story such that the beginning, middle and end are clearly presented in a correct fashion.	3. Story is logical and coherent, with clear structure and good use of language. Spelling and syntax are acceptable, some use of complex sentences and characters in the story are well presented. Demonstrates a command of script.
	2. Able to write a basic story with key features of a narrative but without the structure or style of a competent writer. Uses simple sentence structure and limited words to describe a situation, spelling and syntax are still developing.
1. Low level writing involves a lack of structure to the story, poor spelling and sentence structure; ideas are unconnected. Little reference to the characters in the story or the plot.	1. Uses a basic approach using simple writing skills but with an unclear structure. There are many errors and omissions in the story line. Still showing poor command of script and technical features of writing.

This method of interpretation provided a formal description of the underpinning construct derived from the item and score point descriptions. It also enabled a comparison with the hypothesised construct described in Table 4. Table 8 illustrates the two construct descriptions. The match was close, despite the fact that there were four levels in the derived variable and three in the hypothesised variable. The closeness of the match provided additional evidence that the writing test was measuring a single and identifiable construct.

Setting a Mastery Level

A mastery level defines a level of proficiency or competence at which the student can demonstrably function. In this case the level was defined as the ability to function independently through written language and cope consistently with creative writing tasks. This was determined in a range of ways.

Method 1: Using the variable map

A mastery level can be established using the clusters emerging from the data and illustrated in the variable map. Using the descriptions of the clusters, one cluster or level can be selected as the "mastery level". If, for example, level three were chosen, the scores that a student needed to achieve that level can then be calculated by adding the variable map designations beginning with 1.1 and ending with 9.1. Given

Table 9. *Setting the mastery score by judgement using the variable map*

Item	Max score	Highest possible score below cut point
1	2	1
2	2	1
3	2	1
4	2	1
5	1	1
6	1	1
7	2	1
8	4	3
9	2	1
10	3	1
11	3	1
12	5	2
13	2	0
Σ	31	15

this approach, the score *r* would be the sum of maximum intra-item score codes to the point where the cut line is drawn. In this case, it yields a competency cut score of 15. The score points for each item below the cut level are shown in Table 9.

Method 2: Using a modified Angoff procedure

A second method to set the mastery level uses a modification of the Angoff procedure. Details on this method can be found in Griffin (2001*)*. In this approach, expert judges estimate the proportion of students who represent exact mastery for each possible score (with a minimum category score of zero) on each item (*P(x)*), (where *x*= 0, 1, or 2…). The mean proportion of students whose ability level is at the threshold of mastery and who achieve a specific score is used as the probability that a mastery threshold student would achieve that score. These proportions are then multiplied by the score and the sum of these probabilities represent the minimal cut-off score, which in this case is:

$$r_c = \sum_{i=1}^{k}\sum_{x=0}^{m} x_i . pr(x_i),$$

where r_c is the cut score for mastery, *m* is the score point possible for *k* items, and *x* is the score obtained (*x*=0, …, 5 depending on the item in this case) and *pr(x)* is the probability of a score being obtained or the proportion of students receiving a specific score.

Experienced raters estimated the probability *pr(x)* of a mastery threshold student achieving each score point on each of the thirteen items. The mean value was used as the estimate after deleting outlier estimates. The estimated scores were as shown in Table 10. The proportion of students at the mastery level, *P(x)*, was multiplied by the score category *x*, and these were summed to produce an expected or likelihood score for the item, Σ*x.P(x)*, and these likelihood scores were then summed over all items, ΣΣ*x.P(x)*, to estimate the cut score for mastery based on an idealised student who has just reached the mastery level.

The difference between the cut-off scores calculated by the two methods above is small and remains within the same descriptive level of mastery, as defined by the analysis of the rubrics around the estimates of cut score whether it is 15 or 16 (given that partial scores are not possible). The cut-off point of competency is thus set around the area in Figure 1, at the threshold between level 2 and level 3 where the student moves in ability from a capability to provide a basic structure in the essay to writing a satisfactory conclusion. At this point, the

Table 10. *Mastery for Year 5 Writing (Adapted Angoff)*

Item	p(0)	0*p (0)	p(1)	1*p(1)	p(2)	2*p(2)	p(3)	3*p(3)	p(4)	4*p(4)	p(5)	5*p(5)	Σx*Pr(x)
1	0.1	0	0.3	0.3	0.6	1.2							1.5
2	0	0	0.6	0.6	0.4	0.8							1.4
3	0.1	0	0.6	0.6	0.3	0.6							1.2
4	0.2	0	0.6	0.6	0.2	0.4							1
5	0.3	0	0.7	0.7									0.7
6	0.3	0	0.7	0.5									0.5
7	0.2	0	0.5	0.5	0.3	0.6							1.1
8	0	0	0.4	0.4	0.3	0.6	0.2	0.6	0.1	0.4			2
9	0.2	0	0.6	0.6	0.2	0.4							1
10	0.2	0	0.4	0.4	0.3	0.6	0.1	0.3					1.3
11	0.2	0	0.5	0.5	0.2	0.4	0.1	0.3					1.2
12	0.1	0	0.2	0.2	0.3	0.6	0.2	0.6	0.1	0.4	0.1	0.5	2.3
13	0.7	0	0.2	0.2	0.1	0.2							0.3
												ΣΣx*Pr(x)	15.5

student shifts from writing with errors in both lexis and syntax and begins to express ideas clearly and coherently. Students who have reached the appropriate level for Year 5 transition to Year 6 provide sufficient details to enable visualisation of the characters in their story and have correct Vietnamese script.

The Effect of School Location

Writing ability and skills differed between provinces. Figure 2 reveals large differences in writing ability between the Grade 5 students in the five provinces.

Students from the Ha Noi sample demonstrated a higher level of development than students from the other provinces. This may have been a sampling effect, but it could also be a result of better resources, as reported by Griffin (1998), in terms of both teacher and school quality and the general living standards of Ha Noi compared to those of the other provinces. The average abilities of students in Yen Bai, Thanh Hoa and Vinh Long were below mastery, and at approximately the same level.

An implication is that different teaching strategies are required for different groups of students. At Grade 5 level, average students from Ha Noi were able to write an essay with a clear structure, sufficient detail, and acceptable spelling and syntax. Instruction for these students might best focus on extension of simple language use, and development of imagination and creativity. Although many Quang Nam students had also achieved mastery level, their average ability was lower than that of Ha Noi students. Many still needed instruction and exercises to consolidate vocabulary and

syntax, to incorporate better ideas in their writing, and to improve the cohesion of the whole essay.

Average students in Thanh Hoa, Yen Bai and Vinh Long generally knew the basic features of an essay, but had not yet achieved the structure and style of a competent and creative essay writer. Exercises aimed at these students might be focused on extension of vocabulary, use of complex sentences, and tasks designed to consolidate essay structure and improve cohesion.

Conclusion

A combination of traditional and partial credit scoring rubrics was developed to enable data analysis using the partial credit Rasch model. The variable map presented a clear picture of student distribution and the possible levels of a continuum of creative writing for Grade 5 students in Vietnam. Of the five provinces sampled, Ha Noi students were dominant in their writing performance. Quang Nam was the only other province where the typical student achieved mastery level, established by both the variable map and Angoff method. Students in Yen Bai, Thanh Hoa and Vinh Long were mainly at the second level of the continuum, indicating that they need further, targeted instruction to reach mastery level.

Students who are at different levels of ability require different teaching strategies. Instruction for students in Thanh Hoa, Vinh Long and Yen Bai should be aimed at improving writing structure and style, as well as developing vocabulary and syntax. Quang Nam students would benefit from exercises
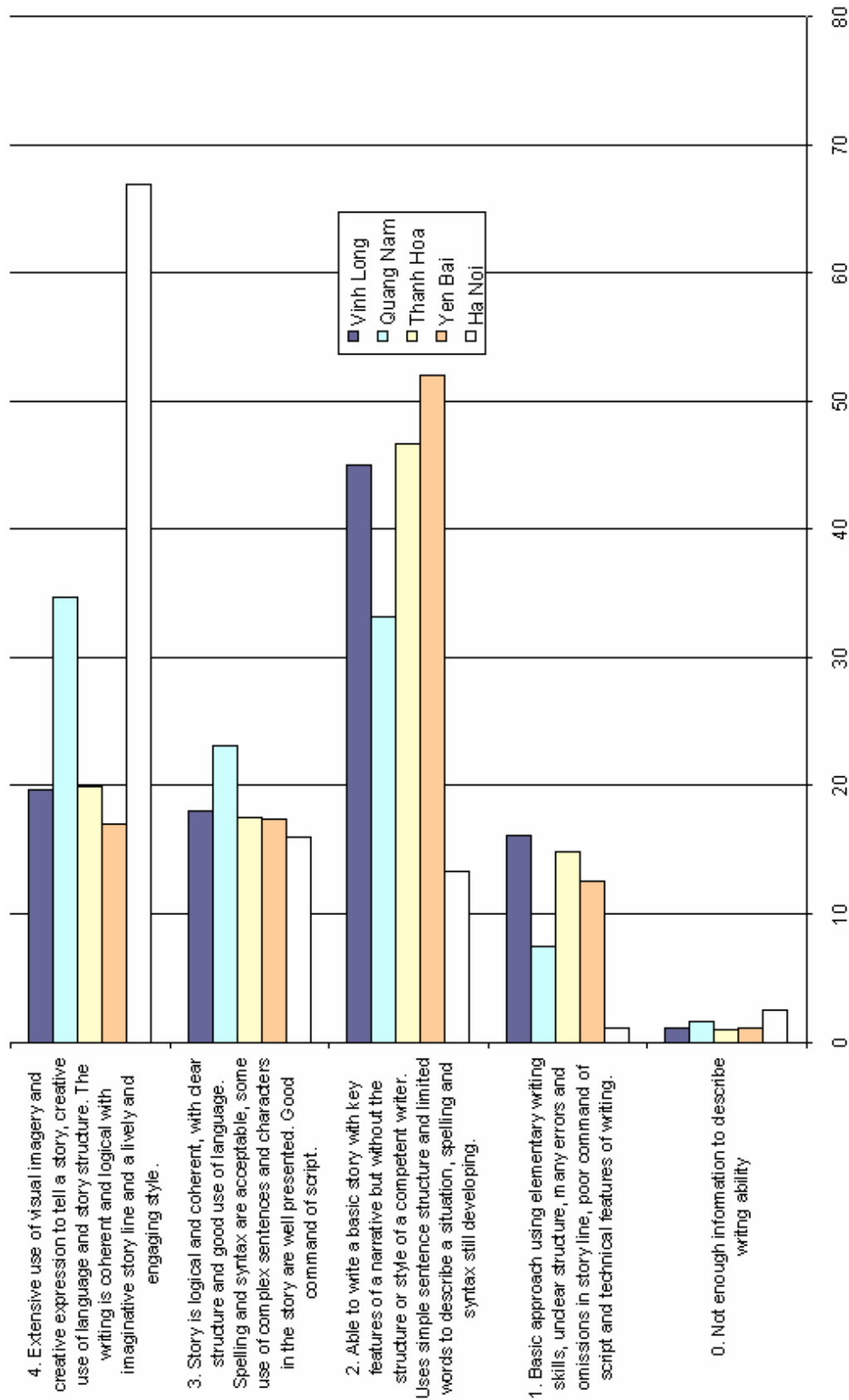
Figure 2. Percentages of pupils at each level of writing competency levels by province

in expanding ideas, employing more complicated language, and developing cohesive writing. Students from Ha Noi, on the other hand, might be given more complex tasks to promote creativity in writing and lively use of language. In addition, variation within provinces indicated that targeted intervention in writing instruction is necessary, perhaps even within schools. This would be most valuable if teachers knew their students' competency levels, and a specific learning plan could be developed for each student.

This study has raised several implications for assessment and marking schemes. The format and design of writing tasks for Vietnamese schools is regulated, leading to consistency. However, description of anticipated levels of writing ability for students at each year level, similar to the levels derived in this study for Grade 5 students, would assist in the design of more appropriate tasks. Similarly, as a method of scoring tests, error counting is less constructive than the approach to scoring rubrics recommended in this study. A detailed scoring rubric can be developed for testing at each grade level to facilitate interpretation of results and assist in curriculum design.

References

Adams, R.J., & Khoo, S.K. (1995). *Quest: Interactive item analysis*. Melbourne ACER.

Ahlawat, K. and V. Billeh (1994). *Achievement of Second Preparatory Students in the Sultanate of Oman on International Tests of Science and Mathematics*. National Centre for Educational Research and Development, Amman.

Ahmed, S. M. and L. T. Salih (1996). *The Assessment of Basic School Children's Learning Achievement in the Sudan*. New York: UNICEF.

Al Nhar, T. (2000). *Regional Report: Learning Achievement of Grade four Elementary Students in Some Arab Countries*. Arab Regional Conference on Education for All: The Year 2000 Assessment (January), Cairo.

Al-Nhar, T. and V. Billeh (1997). *Assessment of Learning Achievement of Grade Students in Oman: Final Report*. Ministry of Education, Sultanate of Oman.

Castro, M. H. (2000). *EFA 2000 Education for All: Evaluation of the Year 2000*. National Institute for Educational Studies and Research, Rio De Janero.

Chinapah, V. (1999). *With Africa for Africa: Towards Quality Eduction for All*. Human Services Research Council, Pretoria, South Africa.

Chinapah, V. (2000). *Equity-Quality Trade-off in Education: Yet many more challenges*. Regional EFA Conference, February 6-8, Warsaw.

Falayajo, W., G. Makoju (1997). *Assessment of Learning Achievement of Primary Four Pupils in Nigeria*, Federal Ministry of Education, Lagos.

Greaney, V., Khandker, S.R., & Alam, M. (1990). *Bangladesh: Assessing basic learning skills*. Bangladesh Development Series. Dhaka: University Press.

Griffin, P. (1990). Profiling literacy development: Monitoring the accumulation of reading skills. *Australian Journal of Education*, 34: 290-311.

Griffin, P. (1998). Vietnamese national study of student achievement in mathematics and Vietnamese. Hanoi, National Institute for Education and Science.

Griffin, P. (2001). Establishing meaningful language test scores for selection and placement. in C. Elder, A. Browne, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, and K. O'Loughlin (Eds.). *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 97-107). Cambridge: Cambridge University Press.

Griffin, P., & Forwood, A. (1990). *Adult literacy and numeracy scales*. Canberra: Department of Employment, Education and Training, Australian Government Printer.

Griffin, P., & Nix, P. (1991). *Assessment and reporting: A new approach*. Sydney: Harcourt Brace Jovanovich.

Khaniya, K. (1999). *National Assessment of Grade 5 Students*. Academy for Educational Development, Katmandu.

Machingaidze, T. and P. Pfukani (1998). *The Quality of Education: Some policy suggestions based on a survey of schools*. Zimbabwe. International Institute for Education Planning, Paris.

Ministry of Education and Training of Vietnam. (1995). Circular 15/GD-DT dated 2 August 1995: Guiding assessment and grading of primary school students (in Vietnamese). Author: Hanoi.

Ministry of Education and Training of Vietnam. (2000a). Notice 481/VP dated 1 June 2000: Conclusion by Deputy Minister Le Vu Hung on adjustment of the primary school Curriculum and Teaching Plan in the year 2000 (in Vietnamese). Author: Hanoi.

Ministry of Education and Training of Vietnam. (2000b). *The curriculum after the year 2000 for nine subjects of primary education* (in Vietnamese). Author: Hanoi.

Ministry of Education and Training of Vietnam. (2001a). *Vietnamese language, Grade 2* (in Vietnamese). Hanoi: Education Publishing House.

Ministry of Education and Training of Vietnam. (2001b).

Vietnamese language, Grade 3 (2 volumes; in Vietnamese). Hanoi: Education Publishing House.

Ministry of Education and Training of Vietnam. (2001c). *Vietnamese language, Grade 4* (2 volumes; in Vietnamese). Hanoi: Education Publishing House.

Ministry of Education and Training of Vietnam. (2001d). *Vietnamese language, Grade 5* (2 volumes; in Vietnamese). Hanoi: Education Publishing House.

Narros, S. and K. A. Mohammed (1998). *The Quality of Education: Some policy suggestions based on a survey of schools. Zanzibar*. International Institute for Education Planning: Paris.

Odeh, D. and M. H. Kizilbash (1998). *Monitoring Learning Outcomes of Class 5 Students. Results of a Mathematics Pilot Survey*. Federal Government and Multi Donor Unit, Lahore, Pakistan.

OECD Program for International Student Assessment, Project Consortium. (2001). *Described proficiency scales for PISA 2000.* Final report to the OED Secretariat.

Rasch, G. (1960, 1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Smith, Jr., E. V. (2001). Reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, *2*, 281-311.

Turner, R. (forthcoming) Chapter 16. Proficiency Scale Construction. In: Adams, R.J. (ed) *PISA 2003 technical Report*. Paris. OECD.

Voigts, F. (1999). The Quality of Education: Some policy suggestions based on a survey of schools. Namibia. UNESCO International Institute for Education Planning, Paris

Wright, B., and Masters, G. (1982). *Rating scale analysis*. Chicago: MESA Press

Wu, M.L. and Adams, R.J. (2005). *Personal communication*.