교육과정평가연구
The Journal of Curriculum and Evaluation
2011, Vol. 14, No. 3, pp. 333~361

# How to Make an Assessment More Informative and Interpretable Using the Ordered Partition Model

Yongsang Lee
(Associate Research Fellow, Korea Institute for Curriculum and Evaluation)

≪ ABSTRACT ≫

As an extension of the partial credit model (PCM), the ordered partition model (Wilson, 1992; OPM) is designed for the measurement context in which different strategies might lead to the same score on an assessment. In the diagnostic context, the data would neither be nominal nor completely ordered, and so it may not be suitable for other polytomous item responses models. The OPM, however, does deal with this type of data (Wilson, 1992). This paper demonstrates how the OPM can help make assessments more informative. To help readers understand the OPM, the PCM and its relationship with the OPM is first described, then, the interpretation of OPM parameters is explained by showing the OPM results with two illustrative data sets: the 'Using Evidence' data and the 'PISA 2003 science assessment' data.

## Ⅰ. Introduction

The nature of assessments can be explained in terms of formative and summative aspects. Summative assessment evaluates how well or how poorly students do in a given subject. In contrast, formative assessment focuses on how students perform, and thus can be used to inform teachers how

they can help students improve their learning. Formative assessment is concerned with how judgments about the quality of student responses can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning (Sadler, 1989). Hence, one way to differentiate formative assessment from summative assessment is in terms of assessment purpose.

In previous research, formative assessment has been approached in terms of its purpose and nature and has been identified by its several elements (Bell & Cowie, 2000; Black & Wiliam, 1998a; Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Tierney & Charland, 2007; Sadler, 1989). Bell and Cowie (2000) described formative assessment based on nine characteristics identified by teachers: (a) responsiveness, (b) sources of evidence, (c) tacit processes, (d) uses of professional knowledge and experience, (e) its importance as an integral part of teaching and learning, (f) formative assessments done by teachers and students, (g) the purposes for formative assessment, (h) the contextualized nature of the process, and (i) the dilemmas. They also point out that teachers identified two main purposes for formative assessment: to inform the students of their learning, and to inform teachers of their teaching. Similarly, Sadler (1989) suggests that feedback is a key element in formative assessment and that its two main audiences are teachers and students. By reviewing recent comprehensive research on formative assessment in a meta-analysis, Tierney and Charland (2007) defined the characteristics of formative assessment as consisting of five elements: (a) clearly delineated learning goals and evaluative criteria; (b) varied approaches that elicit information about learning (including questioning and observation); (c) balanced and descriptive feedback; (d) adjustments following the assessment of teaching and learning methods; and (e) the active participation of students.

The interesting point in this meta-analysis result is that Tierney and Charland (2007) also pay attention to the importance of feedback. They have found that 70% of articles they reviewed have included feedback from students in their results and discussion sections and have concluded that feedback to students and teachers is an essential component of formative assessment. In previous research also, feedback has been discussed as the one of main roles of formative assessment (Bell & Cowie, 2000; Black, 1996; Brookhart, 2004; Sadler, 1989; Tierney & Charland, 2007; William, 2007).

In fact, this feedback can be obtained after applying measurement models to students' response data. At this point, the quality and quantity of the feedback are critically affected by the measurement models one uses. The traditional item response theory (IRT) models (one, two and three parameter logistic models), however, only characterize students in terms of their proficiency or achievement level. The feedback to teachers and students, consequently, would not be as informative or important

as one might expect. Especially in situations in which different students employ different solution strategies, the traditional IRT models are even less satisfactory because they are not designed to capture these solution strategies. The OPM, however, does deal with this type of situation by observing various solution strategies students use under the Rasch model framework.

The Rasch model is a widely applied measurement model designed to calibrate polytomous as well as dichotomous response items. As extensions of the Rasch model, Rasch family models have been developed to fit various measurement contexts focusing on ordinal scale data (Andrich, 1978; Masters, 1982). In some measurement situations, however, as mentioned above, students' responses cannot be completely ordered; for example, individual students with the same level of science literacy might apply different strategies to solving the same science assessment item. If students' responses are analyzed without considering these strategies, much information about their performance may be lost. The ordered partition model (Wilson, 1992), as an extension of the partial credit model (PCM), is designed for this measurement context in which different strategies might lead to the same score on an assessment. In the diagnostic context, the data would neither be nominal nor completely ordered, and so it may not be suitable for other polytomous item responses models.

In this paper, I demonstrate how the OPM can help make assessments more informative. To help readers understand the OPM, I first describe the PCM and its relationship with the OPM; then, I explain how to interpret OPM parameters by showing the OPM results with two illustrative data sets: the 'Using Evidence' data and the 'PISA 2003 science assessment' data.

## Ⅱ. Partial Credit Model

The partial credit model was originally developed by Masters (1982) as an extension of the Rasch Model. It was designed to analyze test items that require multiple steps and for which it is important to assign partial credit for completing several steps in the solution process (Embretson & Reise, 2000). This model allows flexible step difficulties, which means there is no assumption about the relative difficulties of the steps within any item. In the situation in which there are $i$ items ($i$=1,..., I) and in which each item has $m_i+1$ categories graded into $m_i+1$ possible scores $m$ ($x$=0,...,$m_i$) so that $(m_i+1)^{th}$ category has score $m$, person $n$'s probability of reaching performance level $m$ in item $i$ is originally expressed by Masters (1982) as:

$$P(X_{ni}=m)=\frac{\exp(\sum_{j=0}^{m}(\theta_n-\delta_{ij})}{\sum_{h=0}^{m_i}\exp(\sum_{j=0}^{h}(\theta_n-\delta_{ij})} \quad , \quad m=0,\ldots,m_i \tag{1}$$

where $X_{ni}$ is a random variable that represents the response of person *n* with ability $\theta_n$ to item *i*,

$\delta_{ij}$ is an $j^{th}$ level parameter for item *i* associated with score *m,*

$\delta_{i0}$ = 0, so that $\sum_{j=0}^{0}(\theta_n-\delta_{ij})\equiv 0$ , and $\exp\sum_{j=0}^{0}(\theta_n-\delta_{ij})=1$.

The numerator represents the item difficulty of the completed steps, while the denominator is the sum of all possible numerators. Wilson (1992) rewrote this model using Anderson's parameterization (1983) as:

$$P(X_{\ni}=m)=\frac{\exp[m\theta_n-\eta_{im}]}{\sum_{h=0}^{m_i}\exp[h\theta_n-\eta_{ih}]} \tag{2}$$

where $X_{ni}$ is a random variable that represents the response of person *n* with ability $\theta_n$ to item *i*,

$\eta_{im}$ is a level parameter for item *i* associated with score *m*, and

$\delta_{i0}$=0.

Masters' $\delta$ and Anderson's $\eta$ have the following relationship:

$$\sum_{j=0}^{m}\delta_{ij}=\eta_{im}-\eta_{i(m-1)}$$

For the sake of comparison with the OPM parameters, I utilize this parameterization.

# Ⅲ. Ordered Partition Model

As the extension of the PCM, the ordered partition model is useful for situations in which certain responses represent different but equally valued strategies or types of reasoning (Brown, 2004). For the purposes of diagnosis, remediation, and curriculum revision, this model is quite helpful because estimates of how students solve problems could be more valuable than how many they solve (Messick, 1984).

In fact, students' multiple solution strategies can be approached from various modeling perspectives. The OPM is different from these approaches. First, the OPM takes into account the solution strategies of observed students. When their solution strategies, however, cannot be observed, the latent trait model can be applied (Mislevy & Verhelst, 1990; Samejima, 1983). In this framework, it is assumed that individual students belong to one of the mutually exclusive classes that correspond to solution strategies and that the responses from all students in a given class accord with standard IRT models (Mislevy & Verhelst, 1990). Second, the OPM is designed for investigating multiple strategies within items. If each item, however, represents a unique solution strategy, component IRT models can be applied (Embretson, 1985; Butter, De Boeck, & Verhelst, 1998; Smits & De Boeck, 2003).

Once again, to explain the model, let's assume that we have $i$ items ($i$=1,...,I) and each item has $m_i+1$ categories which would be graded into $m_i+1$ possible scores $m$ ($m$=0,...,$m_i$). Person $n$'s probability of reaching performance level $m$ (i.e., score $m$) on item $i$ with a response $k$ could be expressed in the ordered partition model as:

$$P(X_{ni}=k)=\frac{\exp[\theta_n B_i(k)-\xi_{ik}]}{\sum_{h=0}^{m_i}\exp[\theta_n B_i(h)-\xi_{ik}]} \tag{3}$$

where $X_{ni}$ is a random variable that represents the response of person n with ability $\theta_n$ to item $i$,

$\xi_{ik}$ is a level parameter for item $i$ associated with a response $k$, and

$\xi_{i0}=0$.

In this model, $B_i(k)$ is the scoring function for item $i$ in which response $k$ is assigned to level $m$,

so that $B_i(k) = m$. Since the OPM is a generalized model of the PCM, the OPM and PCM parameters can be converted into each other. Equation (4) shows how to convert the Andersen OPM parameters ($\xi_{ik}$) into the Andersen PCM parameters ($\eta_{im}$):

$$\eta_{im} = \ln\left[\frac{\sum_{B_i(k)=m-1} \exp(-\xi_{ik})}{\sum_{B_i(k)=m} \exp(-\xi_{ik})}\right] \quad (4)$$

where $\eta_{im}$ is a level parameter for item *i* associated with score *m*, and $\eta_{i0} = 0$.

More about the OPM and the relationship between parameters are explained later on using the empirical data.

The advantage of the OPM analysis is in testing which scoring scheme best fits the data. In measurement situations, the appropriate scoring scheme is often decided on best theories without much empirical evidence. In the OPM analysis, however, the better scoring scheme can be identified by applying alternatives to given data and by examining model fits to the data (Wilson, 1992; Wilson & Adams, 1993). In order to identify which scoring scheme is better for a given set of data, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Deviance Information Criterion (DIC) can be used. Depending on the sample size and other conditions, these indices do not necessarily agree with each other, and thus the model fit should be carefully investigated to identify the better scoring scheme.

# Ⅳ. Example 1: Conceptual sophistication

## 1. Conceptual Sophistication from "Using Evidence project"

The Center for the Assessment and Evaluation of Student Learning (CAESL) developed an embedded assessment tool to evaluate how students use evidence in their attempt to make a scientific argument. For this purpose, CAESL also built a conceptual framework for modeling students' scientific reasoning based on Toulmin's and Duschl's frameworks (Brown, 2004). In this framework, students' statements would be mapped to the categories of Claim, Premise, Rule, Evidence, or Data, and would

be judged according to a scoring guide; that is, highly advanced students' statements would include every component for the Claim, Premise, Rule, Evidence, and Data, while others might not.

The CAESL framework evaluated the quality of individual scientific argumentation in terms of its Conceptual Sophistication, Precision, Reliability, and Validity. *Conceptual Sophistication* refers to the quality and complexity of the concepts that the Rule implicates, ranging from misconceptions to normative scientific conceptions. *Precision* refers to the degree of specificity to which the Rule is phrased, ranging from ambiguous statements to quantified statements that appropriate units. *Reliability* indicates the quality of the source and the quantity of the data that make up the evidence, ranging from made-up examples to controlled experiments with multiple trials. *Validity* indicates the quality of the reasoning that links one component to another, ranging from no link to valid logical connections. These four variables were developed to represent students' performance level. Each variable has its own steps which are characterized to differentiate students' statements.

As an illustrative example, I have chosen to analyze the 'conceptual sophistication' variable data. In the 'conceptual sophistication' variable, students' progression is represented by seven different ordered levels in order to differentiate students' scientific reasoning from a low level to a high level: (a) the unproductive misconception, (b) the productive misconception, (c) the singular, (d) the relational, (e) the combined, (f) the multi-relational, and (g) the multi-combined level. Since students' performances are characterized by these seven levels of progression, their level of scientific reasoning has been judged referring to these levels. See Figure 1 for how these levels appear for the concept of buoyancy. First, the unproductive misconception (UM) level describes students who use a concept that is incorrect or irrelevant. Second, the productive misconception (PM) level indicates students who use a concept that is incorrect, but could be judged to be relevant in some way. Third, the singular (SI) level characterizes students who use one concept that will later be used in combination with other concepts (i.e., one of mass, volume, or buoyant force). Fourth, the relational (RL) level represents students who use more than one concept that will later be combined, and show that they know a combination is needed (e.g., they know that mass and volume are both needed, but do not yet know how to combine them). Fifth, the combined (CB) level describes students who do combine concepts properly but not completely (i.e., they know what density is and that it is involved). Sixth, the multi-relational (MR) level indicates students who know the combined concept and know how to apply it to the situation, but do not make the final step (i.e., they know that the density of both the object and the liquid is involved, but they do not know what to do with them both). Finally, the multi-combined (MC) level represents students who use the combined concepts correctly and completely (i.e., they know how to calculate and understand relative densities).

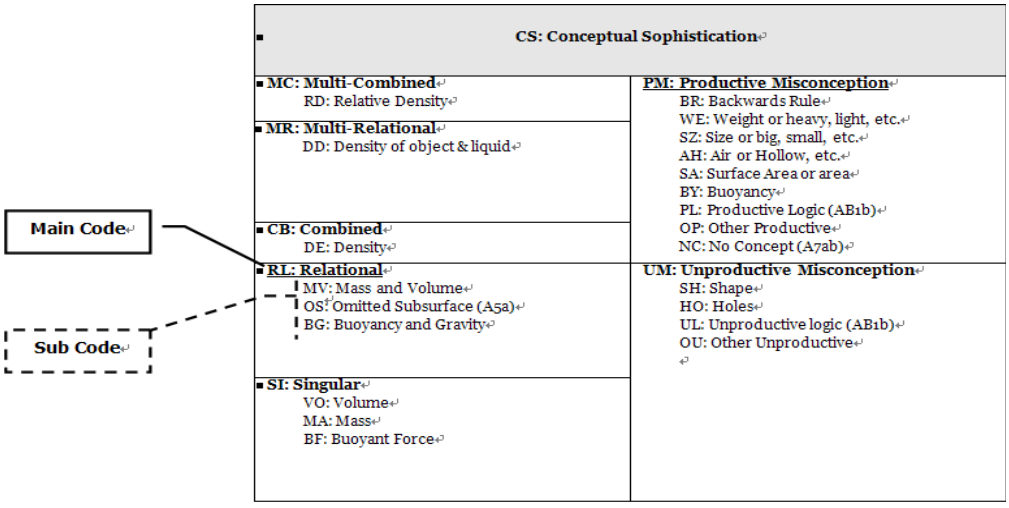
## 2. Data

As a part of this project, WestEd sampled 90 students from an elementary school, 342 students from a middle school and 32 students from a high school for data collection and tried to measure their scientific argumentation ability. During the data collection a certain number of students did not answer for various reasons, and the actual number of students whom they measured is shown in <Table 1>.

〈Table 1〉 Summary of Respondents

| | Elementary | Middle | High | Total |
|---|---|---|---|---|
| Number of respondents | 77 | 234 | 32 | 343 |
| Percentage of overall sample | 86% | 68% | 100% | 74% |

In this data set, there are two sorts of codes: one made using the "main code" scoring guide, and the other made using the "sub-code" scoring guide. In the sub-code scoring guide, each main code has several sub-codes which represent distinct characteristics associated with students' responses. Generally, sub-codes within a main code are not ordered. The main codes and sub-codes for Conceptual sophistication are shown in [Figure 1]




〔FIGURE 1〕 The Scoring Scheme for Conceptual Sophistication

## 3. Results

In this paper, ConQuest (Wu, Adams, & Wilson, 1998) was used to calibrate items using the PCM and the OPM. As ConQuest's parameterization is technically different from what Wilson (1992) discusses about the OPM, however, it would be difficult to interpret parameters. While ConQuest parameterizes the OPM using an item parameter and step parameters, Wilson (1992) parameterizes the OPM using Andersen level parameters (Brown, 2000). In order to interpret OPM parameters, therefore, ConQuest parameters were converted into Anderson level parameters.

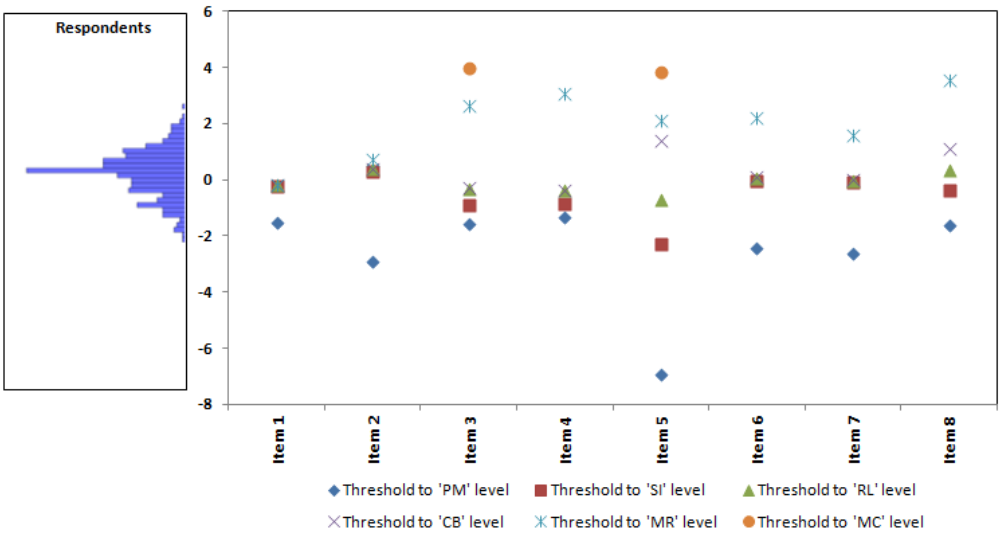
### (1) Partial Credit Analysis

Since ordered performance levels were identified in the responses to items, and since these responses were scored polytomously by multiple raters using the main code scoring scheme, the PCM with rater effects was applied to first calibrate the items. Alphabetical scores were re-coded into a zero-to-six score that respectively ranges from UM (Unproductive Misconception) to MC (Multi-Combined) using ConQuest commands. The PCM results provide an opportunity to observe students' performance levels along with the rater effects on given items.

*Students' Performance Level and The Relationship with Items and Raters*

First, the relationship between students and items can be shown using the Wright map in Figure 2, which maps out the respondents' location and item thresholds. This Wright map shows both the latent distribution of persons and the location of item threshold estimates on the same logit scale. For the sake of convenience only the items scored by rater '1' are presented.

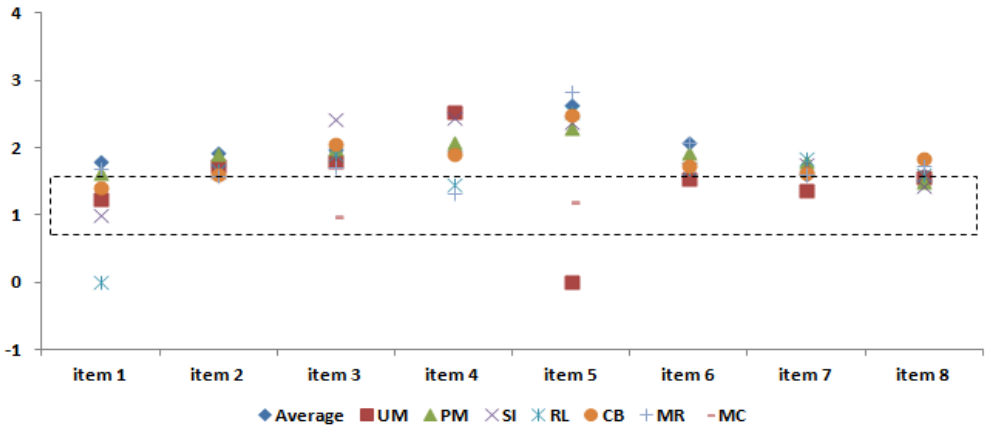
As can be seen in the left side of [Figure 2], respondents are mostly located between logits -2 and 2, and the width of the respondents' location is about 4 logits. On the other hand, the right side of graph shows that most of item threshold parameters are located between logits -4 and 4, which means that the item thresholds cover the entire range of observed levels of conceptual sophistication. These thresholds indicate the proficiency required to achieve a response at that level or above on the item 50% of the time (Kennedy & Wilson, 2007). A threshold of 0 logit for the 'RL' level of Item 8 (a light green triangle), for instance, a means that a student whose location is 0 logit has a 50% chance of providing a response that will be scored in or below the 'RL' level. In this Wright map, item thresholds show us that, for most items, SI, RL, and CB levels can only with difficulty be distinguished and that students might experience almost the same difficulty to achieve these levels.




〔FIGURE 2〕 Wright Map with Thresholds for Conceptual Sophistication

Second, the comparison of the expected number of students responding to the level with observed number responding to that level can be shown using the weighted fit statistics(Wu, Adams & Wilson, 1998) in [Figure 3] These fit statistics are residual based indices and, by convention, we consider them acceptable when they are with the range from 0.75 to 1.33(Adams & Khoo, 1996). The values over 1.33 indicate too much randomness in the level (or score/category), while values under 0.75 indicate less randomness than expected and possibly indicates local dependency.

In [Figure 3] a dotted line box shows the 95% confidence interval of the expected value of fit statistics. As shown in [Figure 3] most of the items and steps are above the range of the confidence interval indicating this data doesn't fit the partial credit model. It also can be interpreted that there were many unexpected responses considering students' proficiency level. Thus, in this conceptual sophistication variable, we would conclude that this model (the PCM) is not fitting, and we should consider alternatives (perhaps alternative items, perhaps alternative models that implicate a theory about the misfit)

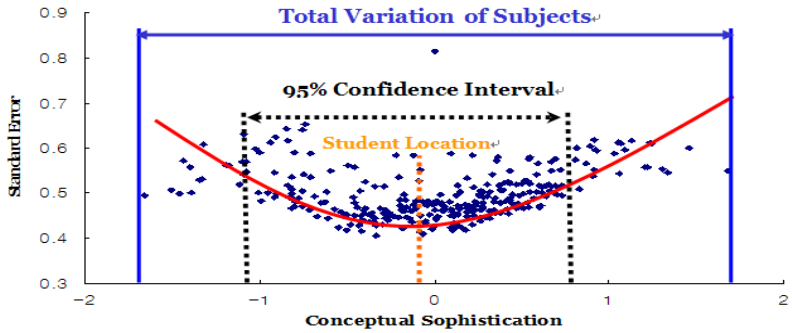



〔FIGURE 3〕 Weighted fit statistics for items of Conceptual Sophistication

*Measurement Error and Reliability*

Since a student's location is only an estimate, there is a certain degree of uncertainty associated with data. This uncertainty is usually characterized using the standard error of the location (known as the standard error of measurement (SEM)), which indicates how accurate each estimate is (Wilson, 2005). [Figure 4] shows the relationship between students' location and the SEM. The absolute value of the SEM for the students seems to be small. When individual students' SEMs are compared with the range of the students' location, however, one notes that these SEMs are actually not that small. For Student 1, for example, whose location estimate is -0.13 logits and standard error is 0.46 logits, the 95% confidence interval of his location is about 1.8 logits wide, about 54% of the width of the students' locations. One could conclude, at this point, that this student's SEM is fairly large and that his results should be interpreted very carefully with the 95% confidence interval in mind.

The closer the respondent is to an item, the more the item can contribute to the estimation of the respondent's location (Wilson, 2005). A smaller SEM indicates that items are nearer to the estimate of the respondent's location. This relationship can also be expressed using the information curve. This information curve displays the most sensitive part of the instrument. Since students located around -0.1 logits show the smallest SEM in general, the assessment used in this project can give the most reliable information for these students. The separation reliability obtained from ConQuest is very high (0.99).

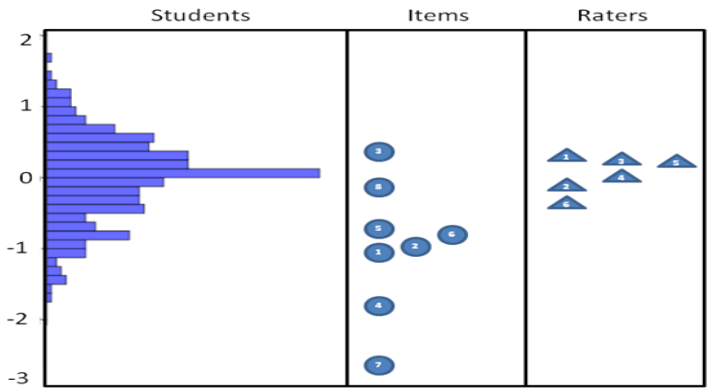



〔FIGURE 4〕 The SEM Curve from the Partial Credit Model

*Rater Effects and Rater Fit*

Since, for this data set, six different raters determined the scores for the students' responses, the raters' levels of harshness might have affected the outcome; that is, the students' scores might not have resulted from simply the combination of the characteristics of the items and those of the students, but might also crucially rely on which rater did the rating. The effect of the rater should thus also be considered when judging the proficiency of the students.

As can be seen in [Figure 5] and <Table 2> the raters' harshness varies approximately from logits -0.18 to 0.2; the range is about 0.38 logits. Since the standard deviation of the latent distribution is about 0.67, estimates of the students' locations may be changed by about 0.5 of the standard deviation of the latent distribution if rater effects are ignored. Furthermore, by considering the standard error of raters' locations shown in <Table 2> it can be seen that rater effects are statistically significant. Rater effects should, therefore, not be ignored when interpreting students' proficiency levels.




〔FIGURE 5〕 The comparison of students, items, and raters' locations

<Table 2> also summarizes the raters' infit statistics and indicates that the raters fit the partial credit model reasonably well, although some are quite low. The low infit value may indicate that the item is locally dependent within some of the raters, but none are below the value of 0.75 mentioned above.

〈Table 2〉 The Raters' harshness and Infit Statistics from the Partial Credit Analysis

| Rater | Rater1 | Rater2 | Rater3 | Rater4 | Rater5 | Rater6 |
|---|---|---|---|---|---|---|
| Rater's harshness | 0.216(0.03) | -0.145(0.03) | 0.206(0.04) | -0.085(0.03) | 0.162(0.03) | -0.354(0.07) |
| MNSQ (weighted) | 0.86 | 0.82 | 1.07 | 0.89 | 1.01 | 0.77 |

( ):Standard error

Thus far, the feedback that the PCM results offer is mainly about the proficiency level and the quality of the assessments of the students. This is, in fact, the main purpose of the PCM analysis. From a formative perspective, however, this feedback is not very helpful because it does not allow a broader understanding on the responses of students to given items. Since this summative feedback does not contain information about which strategy students used to achieve a certain level of proficiency, it is unclear what teachers can do to improve their students' proficiency. For this reason, further analysis using the OPM should be helpful.

### (2) Ordered Partition Analysis

For the OPM analysis, alphabetical sub-codes were also re-coded into a zero-to-twenty-one score that respectively ranges from OU (Other Unproductive) to RD (Relative Density) in the same manner as was done previously in the PCM analysis. Using Equation (3), these numerical sub-codes are mapped into seven different levels: $B_i(0)$=0, ⋯ , $B_i(21)$=6. The rater effects are also considered for the OPM analysis.

As an extension of the PCM, the OPM is able to provide equivalent information to what the PCM provides. Furthermore, it offers more detailed information about students' performances; specifically, how students approach a problem and reach a solution.

*OPM Parameters and Odds*

For the parameter interpretation, since ConQuest provides parameter estimates for the OPM using a

different parameterization than that given by Wilson (1992), ConQuest parameters were converted into Anderson level parameters. Equation (5) shows how Conquest OPM parameters can be converted into Anderson level OPM parameters (Brown, 2004).

$$\eta_{ik} = \xi_i B_i(k) + \sum_{j=0}^{k} \xi_{ij} \tag{5}$$

$\xi_i$ where is ConQuest OPM item parameter for item *i*,

$\xi_{ik}$ is a step parameter for item *i* associated with reaching category *k* from *k*-1,

$\xi_{i0} \equiv 0$, and $\sum_{j=0}^{K_i} \xi_{ij} \equiv 0$ .

For instance, the parameters for Item 1 and Step 1, which indicates the step from Category OU to Category UL, are -1.044 and -0.311, respectively. The Anderson level ordered partition model parameter for Category UL would therefore be calculated as -1.044(0) + 0 + (-0.311) = -0.311.

At the same time, the odds of being in Category NC in the PM level rather than the UM level can be calculated based on Anderson level OPM parameters. Equation (6) and (7) display how the odds can be obtained. With these equations, Wilson (1992) explains how to interpret these parameters.

$$O_{im} = \sum_{B_i(k)} O_{im}^{k} \tag{6}$$

$$O_{im}^{k} = \frac{\exp(\eta_{ik})}{\sum_{B_i(t)=m-1} \exp(\eta)} \tag{7}$$

where $\eta_{ik}$ is Anderson level OPM parameter for item *i* associated with category *k*,

$O_{im}$ is the odds of being in level *m* for item *i,* and

$O_{im}^{k}$ is the odds of being in category *k* in level m rather than level *m*-1.

<Table 3> displays Anderson level parameters and their odds for Item 8. Since there was no response for the top level, <Table 3> shows only levels from UM to MR. As can be seen, the odds of being in subcategories MV (Mass and Volume), OS (Omitted Subsurface), and BG (Buoyancy and Gravity) for the Relational (RL) level are 0.858 to 1, 0.001 to 1, and 0.000 to 1, respectively. One

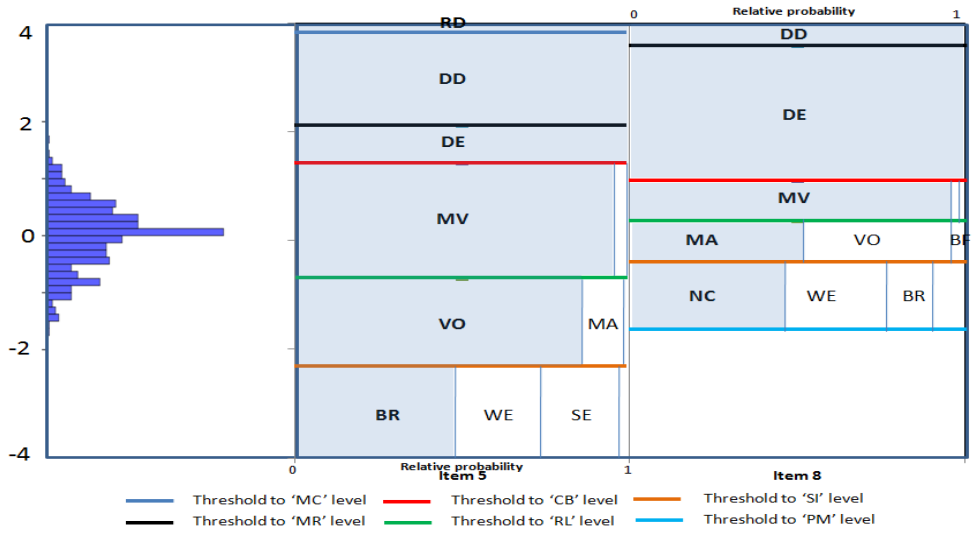
can observe, therefore, that students in the Relational level would most likely be in Category MV. For the Singular level (SI), the odds of being in Category VO (Volume), MA (Mass), and BF (Buoyant Force) are 0.859 to 1, 0.974 to 1, and 0.001 to 1, respectively. It would thus be proper to conclude that students in the Singular level would most likely be in Category MA. In the same way, the odds of being in subcategory NC (No Concept) for the Productive Misconception (PM) level is 2.309 to 1. After comparing the odds of other subcategories in the PM level, one can observe that students in this level are most likely to choose Category NC (see Appendix II for other items).

By mapping these odds in the Wright map, the pattern of odds for each category can be clearly shown and compared to thresholds for levels (e.g., UM, PM, SI, RL, CB, and MR levels). Since thresholds for items are not clearly differentiated (see [Figure 2]), [Figure 6] shows odds only for items 5 and 8 to capture this idea. In [Figure 6] the vertical axis indicates the logit scale for thresholds of items and students' ability; the horizontal axis shows the proportion of odds compared to the total odds of categories within the level. In the 'PM' level of the item 8, for example, the proportion of the odds for the category 'NC' is 0.456 because the sum of odds for this level is 5.057 and the odds for the category 'NC' is 2.309. Thus, compared to the category 'BR' whose proportion of the odds is 0.104, it can be interpreted that students are four times more likely to choose the category 'NC'. This Wright map clearly visualizes this result showing that the category 'NC' is the most dominant category in this level. In contrast, for item 5, the category 'BR' is the most dominant, and it can be interpreted that students in this level are most likely to choose this category. In this Wright map, it can be seen that students adopt one typical solution strategy at a certain level of understanding on the subject; however, it is apparent that this typical strategy is different across items.

〈Table 3〉 Anderson Level OPM and OPM Parameter Odds for Item 8

| Level and Category | OPM parameter | ODDS |
|---|---|---|
| 5 : MR: Multi-Relational | | |
| DD: Density of object & liquid | -0.825 | 0.035 |
| 4 : CB: Combined | | |
| DE: Density | -4.174 | 0.634 |
| 3 : RL: Relational | | |
| MV: Mass and Volume | -4.628 | 0.858 |
| OS: Omitted Subsurface (A5a) | 2.363 | 0.001 |
| BG: Buoyancy and Gravity | 4.446 | 0.000 |

| Level and Category | OPM parameter | ODDS |
|---|---|---|
| 2 : SI: Singular | | |
| VO: Volume | -4.023 | 0.859 |
| MA: Mass | -4.148 | 0.974 |
| BF: Buoyant Force | 2.498 | 0.001 |
| 1 : PM: Productive Misconception | | |
| BR: Backward Rule | -1.998 | 0.573 |
| WE: Weight or heavy, light, etc. | -2.923 | 1.446 |
| SZ: Size or big, small, etc. | -0.62 | 0.145 |
| AH: Air or Hollow, etc. | -0.62 | 0.145 |
| SA: Surface Area or area | -1.719 | 0.434 |
| BY: Buoyancy | 3.012 | 0.004 |
| PL: Productive Logic (AB1b) | 4.765 | 0.001 |
| OP: Other Productive | 5.779 | 0.000 |
| NC: No Concept (A7ab) | -3.391 | 2.309 |
| 0 : UM: Unproductive Misconception | | |
| SH: Shape | -2.473 | |
| HO: Holes | 5.44 | |
| UL: Unproductive logic (AB1b) | 7.751 | |
| OU: Other Unproductive | | |




〔FIGURE 6〕 Item thresholds and OPM odds for items 5 and 8

# Ⅴ. Example 2: the PISA 2003 Science Assessment

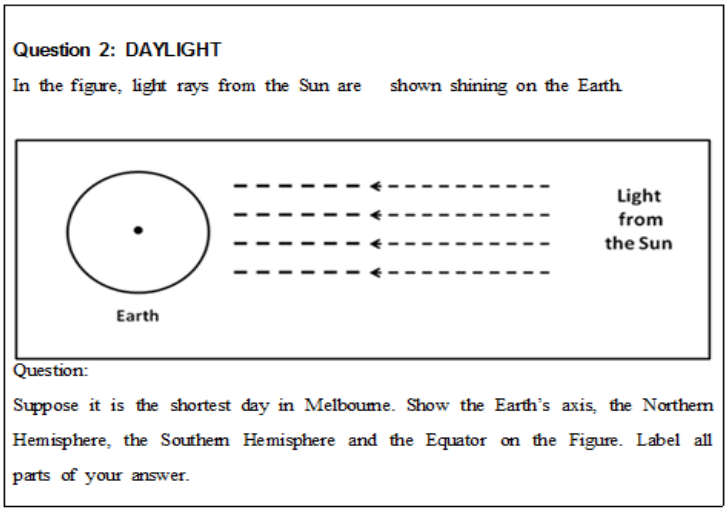
## 1. PISA 2003 Science Assessment

The Programme for International Student Assessment (PISA) is an internationally standardized assessment that was jointly developed by participating countries and was administered to 15-year-olds in schools (OECD, 2003) to identify key demographic, social, economic, and educational factors that affect students' performance in reading, mathematics, and science. For this purpose, fifty-seven counties including Korea participated in 2003. The science section of PISA 2003 included multiple choice (MC), complex multiple choice (CMC), and open constructed response (OCR) items. For OCR items, PISA collapsed the scoring categories to increase the reliability and validity of the science scores across nations. Collapsing these scoring categories however, reduced the amount of information available on the science literacy of students, thus limiting the efficacy of the PISA assessment.

As a second example, I have chosen to analyze the PISA 2003 Science assessment data from Korea.

## 2. Data

The PISA 2003 assessments were administered to between 4,500 and 10,000 students who were fifteen years old (10th grade) in each participating country (OECD, 2003). For this second example, the original coding of eight OCR items from ACER were obtained, of which five were selected for OPM analysis. In total, 5,444 students in Korea took the PISA 2003 science assessment, and 1,681 students responded to these five OCR items. In the original PISA 2003 data set, OCR items were coded using scoring guides consisting of item-specific categories; for example, for Item 'S129Q02' (Figure 7), student responses were coded with nine codes as shown in <Table 4> As can be seen in <Table 4> each code represents unique response characteristics, but PISA scored the item using only three: full credit (2), partial credit (1), and no credit (0). As a result, the same score was assigned to more than one code. The publicly available PISA 2003 data set includes only these scores. For illustrative purposes, I discuss results for only Item "S129Q02".




〔FIGURE 7〕 PISA 2003 Science Item 'S129Q02'

〈Table 4〉 The Original Codes for Item 'S129Q02'

| Score | Code | Description |
|---|---|---|
| Full credit | 21 | Diagram with Equator tilted towards the Sun at an angle between 10° and 45°and Earth's axis tilted towards the Sun within the range 10° and 45° from vertical, and the Northern and or Southern Hemispheres correctly labelled (or one only labelled, the other implied). |
| Partial credit | 13 | Angle of tilt of Equator between 10° and 45°, and angle of tilt of axis between 10° and 45°, but Northern and Southern Hemispheres not correctly labelled (or one only labelled, the other implied, or both missing). |
| | 12 | Angle of tilt of Equator between 10° and 45°, Northern and/or Southern Hemispheres correctly labelled (or one only labelled, the other implied), but angle of tilt of axis not between 10° and 45°; or axis missing. |
| | 11 | Angle of tilt of axis between 10° and 45°, Northern and/or Southern Hemispheres correctly labelled (or one only labelled, the other implied), but angle of tilt of Equator not between 10° and 45°; or Equator missing. |
| No credit | 4 | No features are correct, or other responses |
| | 3 | Angle of tilt of axis between 10° and 45° is the only correct feature |
| | 2 | Angle of tilt of Equator between 10° and 45° is the only correct feature. |
| | 1 | Northern and or Southern Hemispheres correctly labelled (or one only, the other implied) is the only correct feature. |
| Missing | 99 | missing |

## 3. Results

Since PISA adopted the reduced scoring categories (no credit, partial credit, and full credit) for data analyses, the PCM has been applied to this data first, and then the OPM analysis has been implemented later to compare with the PCM analysis results. Once again, the OPM analysis results show how to handle nominal categories within each ordinal level, and illustrate what benefit and/or information can be taken from the OPM analysis compared to the PCM analysis.

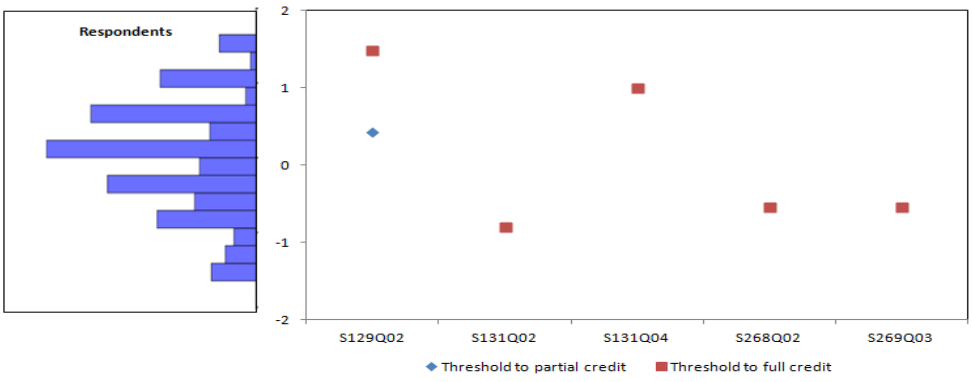
### (1) Partial Credit Analysis

The partial credit analysis results show that the model fit data for this item very well. The infit statistics for the item and step parameters of item 'S129Q02' are almost 1 as shown in Table 5, and other items have essentially the same infit statistics (0.97 ~ 1.7).

〈Table 5〉 Item and step parameters and their fit statistics for the item 'S129Q02'

| S129Q02 | Item | step to partial credit | step to full credit |
|---|---|---|---|
| Estimate | 0.954 | -0.098 | 0.098 |
| MNSQ(Infit) | 1.02 | 0.99 | 1.00 |

Once again, the relationship between the students' proficiency and the items' difficulty can be shown using the Wright map. [Figure 8] maps out students' proficiency and item thresholds in the same logit scale.




〔FIGURE 8〕 Wright Map with Thresholds for PISA 2003 Science items

As shown in [Figure 8] for the item ‘S129Q02’, the threshold of partial credit is clearly different from the full credit. This result indicates that students in the full credit level are quantitatively distinguishable from those in the partial credit level. Since it doesn’t provide any detailed information differentiating students besides their science literacy levels, however, teachers might not be able to use this information to modify the science curriculum and their teaching to improve their students’ science literacy. In order to make the PISA assessment more informative, the measurement model should consider any qualitative differences in students’ responses to given items. Since the PISA 2003 data set includes original codes representing the unique characteristics of responses as shown in Table 4, it is possible to capture qualitative differences in students’ responses by using the OPM model, and thus an OPM analysis needs to be done.

### (2) Ordered Partition Analysis

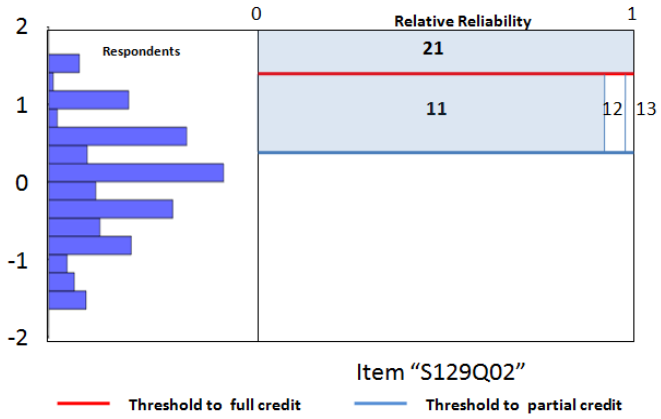
By applying the OPM to the PISA 2003 science data, the science literacy of the students can be investigated in terms of response characteristics as well as their achieved levels. For Item ‘S129Q02’, the PCM analysis results display that students can be quantitatively differentiated in terms of their achieved levels. In the OPM analysis, qualitatively different responses in the same level have been investigated using OPM parameters and their odds. <Table 6> shows OPM parameters and their odds for Item ‘S129Q02’.

〈Table 6〉 Anderson Level OPM and OPM Parameter Odds for Item ‘S129Q02’

| Score | Code | OPM parameter | ODDS |
|---|---|---|---|
| Full Credit | 21 | 6.837 | 0.0016 |
| Partial Credit | 13 | 4.732 | 0.0054 |
| | 12 | 2.717 | 0.0408 |
| | 11 | 0.491 | 0.3783 |
| No credit | 4 | 0.604 | |
| | 3 | 2.679 | |
| | 2 | 6.034 | |
| | 1 | 0 | |

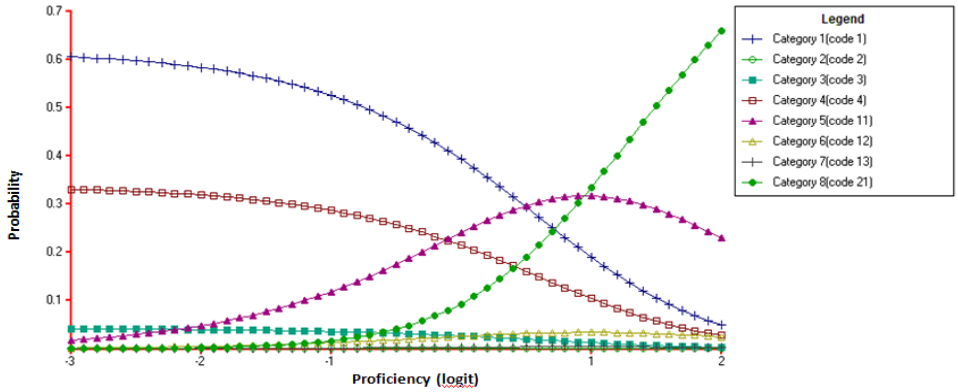
As can be observed in bold, the odds for code 11, that is, responses that a) state the angle of the tilt of axis as being between 10° and 45° and b) label correctly the Northern and/or Southern

Hemispheres, but c) state incorrectly that the angle of tilt of the Equator is not between 10° and 45°, or d) miss the Equator is 0.3783 to 1. Compared to other responses, it can be seen that this response is the most probable one for a partial credit. This result can be visualized by incorporating the odds into the Wright map as well. As shown in [Figure 9] the proportions of odds indicate which type of response is the most probable one for a certain level. As observed in this Figure, students for the partial credit level are most likely to give certain types of responses which are coded as '11'.




〔FIGURE 9〕 Item thresholds and OPM odds for item 'S129Q02'

Furthermore, one can observe the category probability curves in [Figure 10] and conclude that this category works for identifying students in the middle level of proficiency. The figure also provides information about the students' learning process.




〔FIGURE 10〕 Category probability curves for item 'S129Q02'

Also noticeable in this figure is that three types of responses have the highest probability for a certain range of proficiency. In the lower level of proficiency, students are most likely able to only label the Northern and/or Southern Hemispheres correctly (Code 1); this answer is treated as an incorrect response. On the other hand, the responses of proficient students not only include the correct labeling of the Northern and/or Southern Hemispheres, but also include a diagram with the Equator tilted towards the Sun at an angle between 10° and 45° and the Earth's axis tilted towards the Sun within the range 10° and 45° from the vertical (Code 21). Between these two groups there is another group of students whose response includes an angle of tilt of axis between 10° and 45° and correctly labels the Hemispheres but does not have an angle of tilt of the Equator between 10° and 45° or has the Equator missing (Code 11). This group shows the learning progress of students based on the concept assessed by Item 'S129Q02'. Based on this result, it can be hypothesized that students might need to first understand the Sun's relationship to the Earth's axis in order to respond to this problem correctly. This information about the learning progression of the scientific concepts that students have offers teachers ideas about how to improve their students' learning and/or their methods of teaching.

# Ⅵ. Discussion

As an extension of the PCM, we have shown that the OPM is quite useful in several ways. From the measurement perspective, the OPM provides a framework to deal with ordinal and nominal codings simultaneously. In measurement situations, the assessments often employ nominal as well as ordinal coding to catch strategies students use to achieve solutions. Traditional IRT models, however, cannot deal with this mixed coding, and so the nominal coding has often been ignored in previous analyses. For formative assessment, the OPM can provide more detailed feedback to teachers. While traditional IRT models can inform only students' achievement (or proficiency) levels, the OPM makes it possible to also see how students approach given items as well as how many items they solve. Since feedback is one of the key elements in formative assessment, the OPM can contribute to formative assessment by upgrading the feedback both qualitatively and quantitatively. Note that the application of the OPM requires to meet a couple of conditions; 1) The scoring scheme should be designed to specify students' strategies as many as possible. 2) The coding process should happen in two ways; the main coding and sub-coding. These two conditions are, in fact, quite challenging to general researchers and teachers. In addition, the interpretation of the OPM model parameters is not

easy at all. For these reasons, the application of the OPM model is somewhat limited. Thus, a researcher may want to apply a simpler model (e.g., PCM) first, and then for the further investigation, he or she can apply the OPM model. The process of the OPM application can be summarized as follows:

a. Apply the simpler model (e.g., PCM) to data.
b. Apply the OPM models.
c. Compare the model fit values to identify the best scoring scheme for a given data set.
d. Convert the OPM parameters to the odds for the interpretation.
e. Mapping these odds in the Wright map and investigate the pattern of odds to identify the best solution strategy to get a certain level of achievement.

These illustrative analyses show what kinds of information the OPM can provide and how this information helps one to understand the performance of students. The first example demonstrates the relationship between students' proficiency levels (main code) and their response characteristics (sub-code). Although the comparison of the PCM and the OPM, the analysis results show that the PCM still fit this data better than the OPM (The AIC values for the OPM and the PCM are 9494.07 and 8293.95, respectively), the OPM may be more beneficial to teachers. By understanding this relationship, teachers may be able to postulate how to improve their proficiency on a given concept. The second example demonstrates how the OPM allows deeper insight into the knowledge and the learning progression of students. If we had collapsed these categories in the PISA 2003 science data, important information, and consequently, the opportunity to refine students' learning progression, would be lost. It is also true in the national-level achievement test which is conducted by KICE (Korea Institute for Curriculum and Evaluation). Since this test has open response items, it is possible to identify students' solution strategies using the OPM model. Thus, the application of the OPM model in the national-level achievement test will be beneficial and should be done.

In reality, the solution strategies students choose vary depending on items. When items represent multiple constructs and the same solution strategies are still applicable across items, it is conceptually possible to construct a multidimensional OPM. At the same time, it might be interesting to apply an ordered partition scoring scheme in the MIRID framework if each item (component item) were to reflect a unique concept and if item parameters to predict other item parameters (composite item). By doing so, one would be able to compare component items in terms of how students respond or solve each component item, allowing one to look at students' reactions to each concept as well as the relationship between concepts.

# References

Adams, R. J., & Khoo, S.T. (1996). Quest. Melbourne, Australian Council for Educational Research.

Andersen, E. B. (1983). A general latent structure model for contingency table data. In H. Wainer & S, Messick (Eds), *Principals of modern psychological measurement* (pp. 117-138). Hillsdale NJ: Erlbaum.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometirka, 43*(4), 561-578.

Bell, B., & Cowie, B. (2000). The Characteristics of Formative Assessment in Science Education. *Assessment in Education, 6,* 101-115.

Black, P. (1996). Formative assessment and the improvement of learning. *British Journal of Special Education, 23*(2), 51-56.

Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record, 106*(3), 429-458.

Brown, N. J. S. (2004). Interpreting ordered partition model parameters from ConQuest. University of California, Berkeley Evaluation & Assessment Research Center.

Brown, N. J. S. (2004). Proposal for Using Evidence project: An Analysis of US and German Science Teaching and Learning.

Embretson, S. E., & Reise, S. P. (2000). Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.

Herman, J. L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). The nature and impact of teachers' formative assessment practices. CSE Technical report 703.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement, 21*(3), 215-237.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*(2), 195-215.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Sadler, D. R. (1989). Formative assessment: revisiting the territory. *Assessment in Education, 5*(1), 77-84.

Tierney, R., & Charland, J. (2007). Stocks and prospects: Research on formative assessment in secondary classrooms. Paper presented in the annual meeting of AERA, Chicago, IL.

William, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 1051-1098). Reston, VA: National Council of Teachers of Mathematics.

Wilson, M. (1992). The ordered partition model: An extension of the partial credit model. Applied *Psychological Measurement, 16*(4), 309-325.

Wilson, M., & Adams, R. J. (1993). Marginal maximum likelihood estimation for the ordered partition model. *Journal of Education Statistics, 18*(1), 69-90.

Wu, M. L., Adams, R. J., & Wilson, M. (1998). ACER ConQuest user guide. Hawthorn, Australia: ACER Press.

· 논문접수 : 2011-09-01/ 수정본접수 : 2011-10-10/ 개재승인 : 2011-10-25

# 초록

## 평가의 환류 및 해석을 위한 OPM 모형 활용 방안

이용상(한국교육과정평가원 부연구위원)

문항 반응 모형들 중에 하나로서 Ordered Partition Model (OPM)은 명목 척도와 서열 척도가 혼재되어 있는 자료를 이용하여 학생들이 사용하는 문제 해결 전략에 대한 정보를 제공하기 위해 개발되어졌다(Wilson, 1992). 본 연구에서는 명목 척도와 서열 척도를 동시에 고려하는 OPM (Ordered Partition Model) 모형을 이용하여 학생들의 문항 반응 자료를 분석하고 해석하는 방법을 소개하고, 명목 척도를 고려하지 않는 다른 모형(PCM)과 비교하여, OPM 모형이 가지는 장점과 한계점을 논한다. 더불어 'PISA 2003' 자료와 'Using Evidence' 자료를 이용하여 OPM 모형이 어떠한 방식으로 학생들의 문제 해결 능력에 대한 보다 풍부한 정보를 제공하는지를 설명하고, 향후 OPM 모형을 적용할 수 있는 연구 영역에 대한 제언을 한다.

*주제어 : OPM, 부분 점수 모형, PISA 과학영역 검사 도구*

## Appendix I

Anderson OPM parameters can be converted into Anderson PCM parameters using Equation (4). For example, this equation gives the Anderson Level 3 (Relational level) PCM parameter for Item 1:

Anderson Level PCM Parameters for Item 1

| Level | 0(UM) | 1(PM) | 2(SI) | 3(RL) | 4(CB) | 5(MR) | 6(MC) |
|---|---|---|---|---|---|---|---|
| Parameter estimate | 0 | -2.284 | 2.439 | 5.385 | -6.690 | -1.044 | -3.209 |

| | item5 | | item6 | | item7 | | item8 | |
|---|---|---|---|---|---|---|---|---|
| | OPM | ODDS | OPM | ODDS | OPM | ODDS | OPM | ODDS |
| **6 : MC: Multi-Combined** | | | | | | | | |
| RD: Relative Density | -4.536 | 0.029 | | | | | | |
| **5: MR: Multi-Relational** | | | | | | | | |
| DD: Density of object & liquid | -8.082 | 0.232 | -4.06 | 0.152 | -14.065 | 0.284 | -0.825 | 0.035 |
| **4: CB: Combined** | | | | | | | | |
| DE: Density | -9.543 | 0.188 | -5.941 | 10.140 | -15.324 | 16.488 | -4.174 | 0.634 |
| **3: RL: Relational** | | | | | | | | |
| MV: Mass and Volume | -11.172 | 2.997 | -3.507 | 1.588 | -12.298 | 0.989 | -4.628 | 0.858 |
| OS: Omitted Subsurface (A5a) | -8.07 | 0.135 | 1.754 | 0.008 | -7.668 | 0.010 | 2.363 | 0.001 |
| BG: Buoyancy and Gravity | -3.207 | 0.001 | -1.383 | 0.190 | -10.873 | 0.238 | 4.446 | 0.000 |
| **2: SI: Singular** | | | | | | | | |
| VO: Volume | -9.929 | 10.740 | -1.247 | 0.034 | -11.79 | 0.163 | -4.023 | 0.859 |
| MA: Mass | -8.074 | 1.680 | -2.855 | 0.171 | -11.387 | 0.109 | -4.148 | 0.974 |
| BF: Buoyant Force | -1.383 | 0.002 | 1.934 | 0.001 | -7.417 | 0.002 | 2.498 | 0.001 |
| **1: PM: Productive Misconception** | | | | | | | | |
| BR: Backwards Rule | -6.765 | 866.862 | -3.836 | 23.071 | -12.019 | 51015.560 | -1.998 | 0.573 |
| WE: Weight or heavy, light, etc. | -6.257 | 521.589 | -3.096 | 11.007 | -12.315 | 68588.898 | -2.923 | 1.446 |
| SZ: Size or big, small, etc. | 0.259 | 0.772 | -0.793 | 1.100 | -6.903 | 306.092 | -0.62 | 0.145 |
| AH: Air or Hollow, etc. | 2.613 | 0.073 | -1.892 | 3.302 | -10.197 | 8249.319 | -0.62 | 0.145 |
| SA: Surface Area or area | -6.255 | 520.547 | 2.678 | 0.034 | -6.941 | 317.947 | -1.719 | 0.434 |
| BY: Buoyancy | 1.628 | 0.196 | -2.987 | 9.871 | -12.871 | 119597.350 | 3.012 | 0.004 |
| PL: Productive Logic (AB1b) | 4.696 | 0.009 | 2.941 | 0.026 | -6.574 | 220.277 | 4.765 | 0.001 |
| OP: Other Productive | 7.485 | 0.001 | -1.475 | 2.176 | -4.565 | 29.544 | 5.779 | 0.000 |
| NC: No Concept (A7ab) | 10.218 | 0.000 | 3.009 | 0.025 | -3.466 | 9.844 | -3.391 | 2.309 |
| **0: UM: Unproductive Misconception** | | | | | | | | |
| SH: Shape | 13.481 | | 5.533 | | -0.086 | | -2.473 | |
| HO: Holes | 15.947 | | 0.018 | | 0.37 | | 5.44 | |
| UL: Unproductive logic (AB1b) | 9.046 | | 3.796 | | 0.753 | | 7.751 | |
| OU: Other Unproductive | 0 | | 0 | | 0 | | 0 | |

# Appendix II

| | item1 | | item2 | | item3 | | item4 | |
|---|---|---|---|---|---|---|---|---|
| | OPM | ODDS | OPM | ODDS | OPM | ODDS | OPM | ODDS |
| **6: MC: Multi-Combined** | | | | | | | | |
| RD: Relative Density | | | | | 1.668 | 0.029 | | |
| **5: MR: Multi-Relational** | | | | | | | | |
| DD: Density of object & liquid | -5.22 | 24.754 | -4.935 | 1.141 | -1.873 | 0.062 | -8.78 | 0.057 |
| **4: CB: Combined** | | | | | | | | |
| DE: Density | -2.011 | 803.948 | -4.803 | 47.572 | -4.658 | 25.437 | -11.638 | 35.061 |
| **3: RL: Relational** | | | | | | | | |
| MV: Mass and Volume | 4.864 | 0.004 | -0.882 | 0.045 | -1.297 | 0.171 | -7.948 | 0.167 |
| OS: Omitted Subsurface (A5a) | 6.76 | 0.001 | 2.148 | 0.002 | 3.67 | 0.001 | -3.421 | 0.002 |
| BG: Buoyancy and Gravity | 7.79 | 0.000 | 3.525 | 0.001 | 0.775 | 0.022 | -5.918 | 0.022 |
| **2: SI: Singular** | | | | | | | | |
| VO: Volume | 9.202 | 0.000 | -1.549 | 0.012 | -0.663 | 0.420 | -7.674 | 0.586 |
| MA: Mass | -0.418 | 0.065 | -3.832 | 0.120 | -2.965 | 4.199 | -9.603 | 4.031 |
| BF: Buoyant Force | 0.676 | 0.022 | -0.888 | 0.006 | 4.561 | 0.002 | -2.737 | 0.004 |
| **1: PM: Productive Misconception** | | | | | | | | |
| BR: Backwards Rule | -1.402 | 1.717 | -0.63 | 0.152 | 7.242 | 0.001 | -5.535 | 0.186 |
| WE: Weight or heavy, light, etc. | -2.675 | 6.132 | -4.673 | 8.639 | -1.158 | 3.184 | -7.65 | 1.543 |
| SZ: Size or big, small, etc. | 4.125 | 0.007 | -0.63 | 0.152 | -0.278 | 1.320 | -6.823 | 0.675 |
| AH: Air or Hollow, etc. | 6.182 | 0.001 | -5.505 | 19.851 | 5.89 | 0.003 | -4.877 | 0.096 |
| SA: Surface Area or area | 7.281 | 0.000 | -2.827 | 1.364 | 7.92 | 0.000 | -5.57 | 0.193 |
| BY: Buoyancy | 0.16 | 0.360 | -2.239 | 0.757 | 2.236 | 0.107 | -1.724 | 0.004 |
| PL: Productive Logic (AB1b) | -1.161 | 1.349 | 1.958 | 0.011 | 5.832 | 0.003 | -0.139 | 0.001 |
| OP: Other Productive | 0.544 | 0.245 | 3.623 | 0.002 | 7.42 | 0.001 | 0.817 | 0.000 |
| NC: No Concept (A7ab) | 4.175 | 0.006 | 4.614 | 0.001 | 8.47 | 0.000 | 1.486 | 0.000 |
| **0: UM: Unproductive Misconception** | | | | | | | | |
| SH: Shape | 6.633 | | -2.105 | | 0 | | -7.215 | |
| HO: Holes | 7.514 | | -1.136 | | | | 0.053 | |
| UL: Unproductive logic (AB1b) | -0.311 | | 2.707 | | | | 2.201 | |
| OU: Other Unproductive | 0 | | | | | | 0 | |