Nonparametric Item Response Theory

W. Blaine Dowler

May 28, 2011

1 Limitations of Parametric Item Response Theory

When assessment items are modeled with a Rasch model, certain assumptions are present which do not always apply. One such assumption is that the likelihood of responding correctly to a particular item increases in a smoothly defined fashion. This is not always the case. In some cases, as students are exposed to related skills, they rapidly assimilate previously presented information, causing a spike in difficulty. A second assumption is that, at some point, every student will respond to a question correctly. This is not necessarily the case. If the question is to multiply a 30 digit number by a 30 digit number without the aid of a calculator, odds are excellent that any given individual will make a computational error at some point in the process (which involves a minimum of 1741 distinct operations, not counting regrouping) and get the item incorrect. Finally, it assumes that the odds of correctly answering a question are strictly increasing with student ability. While this is usually true in the classroom situation, there are exceptions[1] to this rule. When item response theory is applied to political opinion polls and other non-classroom situations, these situations become far more common.

[1]The author's favorite exception: "given that $x + y = 2$ and $xy = 3$, what is $\frac{1}{x} + \frac{1}{y}$? High school students will attempt to solve this through substitution and typically get stuck when they reach an equation which can only be solved by taking the square root of a negative number. Junior high students haven't seen that method yet, and solve the problem by looking for a common denominator and finding that $\frac{1}{x} + \frac{1}{y} = \frac{x+y}{xy} = \frac{2}{3}$.

2 Benefits and Limitations of Nonparametric Item Response Theory

Nonparametric item response theory concerns itself only with relative relative performance of the individuals being assessed. No absolute ability or difficulty scales are used at all. Instead, examinees are sorted and ranked by ability and ability alone. When it is done, you may know that Alice has more ability than Bob, that Carol comes in third and that Ted is at the bottom of the pile, but you will not know the degree to which the differences arise. Because of these limitations, nonparametric item response theory has very limited applications to academia. It may be used to screen applicants for a private institution of some sort, but is nearly useless when measuring student performance in any other respect. It is, however, frequently applied outside academia.

Most opinion polls that are analyzed by professional statisticians are based on this theory. The analysis is performed almost exclusively through variance and covariance calculations from statistics, and the methodologies lend themselves nicely to questions that fall on a rating scale, such as "strongly agree, agree, neutral, disagree or strongly disagree" with each possible response being assigned a numeric value. In this type of analysis, conversion of responses may be necessary. For example, if left wing politics lead to a level of agreement with item 1 but disagreement with item 2, then the values of one of these items should be reversed so that the numeric values assigned are comparable. Similarly, items that are regarded with the same importance are to be treated equally, they must have the same extremes for numeric values. For example, if the five options above are valued such that "strongly agree" is assigned the value 2 and "strongly disagree" is assigned -2, then another item on the same survey with only "yes" or "no" options available must be assigned with values of 2 and -2 (in some appropriate order) to be equitably compared to the earlier item.

There is one other benefit to nonparametric item response theory which can be useful in the analysis of long assessments. The larger the assessment, the more likely that it is multidimensional, meaning it will assess multiple skills at once. Students who are more adept at one group of skills than another can have seemingly inconsistent results with parameterized item response theory. The prevalence of covariance measures in the nonparametric version makes analysis of correlations simple. The theory can identify which assessment items measure the same skill or group of related skills, and which belong to a different group. Armed with this information, traditional parameterized item response theory can be applied along two distinct lines, producing two distinct scores in the different skill groups. As such, it can be an effective companion to a parameterized analysis of an assessment tool, but still does not serve as an effective academic tool in isolation.

3 Final Recommendations

Given the variety of assessment tools and purposes available, recommendations will depend to a degree on the intended purpose of the assessment. The author's recommendations are as follows:

Goal	**Recommendation**
To give a general picture of student achievement within a course or curriculum.	A letter grade serves this purpose best.
To inform students of their strengths and weaknesses. (This should be a goal of every academic course.)	A skills checklist derived from summative assessments, by a mile. This gives students exactly what they need to improve in the long term. A letter grade, percentage, or norm-referenced mark of some form may be a useful supplement, but does not provide enough information on its own to reach the goal.
To screen candidates for acceptance to a job or academic program.	A norm referenced assessment. If space is limited, use a nonparametric item response theory analysis. If space is plentiful, use a parameterized analysis.
To evaluate the effectiveness of instruction.	Norm referenced analysis of criterion referenced assessments. Compare the average student performance to the actual curricular outcomes, with particular attention to any frequent results on the students' skill checklists.

4 Conclusion

The world of assessment is changing right now. The traditional methods of running with a teacher's gut instincts are being overturned by methodologies that are based on research and mathematical methods to best determine what abilities students have and how to improve them. The rate at which this change will occur will depend greatly on the funding provided to education in the future, and that will depend on the voiced opinions of voting taxpayers. Speak your mind to the local politicians about education (and every other issue that concerns you), and vote for the one who listens best.