The Future of Measurement: Making learning visible

John Hattie
Melbourne Education Research Institute (MERI)
University of Melbourne
Chinese Han Dynasty (206 BCE to 220 CE)

- Entrance examination for the civil service from observation & testimonials to written examinations

To avoid nepotism and corruption.
Classical model 1904+

- Presence of errors in measurement
- True score = Observed + Error

Lord and Novick (1968)

\[ X_{jg} = t_{jg} + E_{jg} \]

It is simple, elegant, and surprisingly durable
Item Response Theories  1950s+

\[ p(u=1 \mid \theta) = c + (1-c) \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}} \]

b  Difficulty

a  Discrimination

c  Guessing

Major advances

- test-score equating
- detection of item bias
- adaptive testing
- extensions to polytomous items, to attitudes and personality
Moving beyond classical IRT 2000’s+

Newer models based on:

• time completing items
• multiple attempts on items
• multiple abilities or cognitive processes
• multidimensionality

Advances

• computer based testing of many forms
• development of reporting engines
• computerized essay scoring
• beginnings of integrating of cognitive processes
• merging testing advances with instruction
Advances in Technology

1870’s

2000’s
Advances in teaching?

Teaching has not changed (much) over the past 200 years
Assessment in classrooms

Backward Design
- % of students talked to
- student questions
- peer feedback
- collaboration among students
- early years of reading
Software Applications

Interactive video

Hypermedia testing

Learner controlled
Interactive solutions

Virtual Realities
Reading self-efficacy scale

"...What about you, do you think you're a good reader?"

- Yes, I do
- No, not really
Preference 72% Web 28% Paper

Test Information Functions for Paper-and-Pencil and Web-based Versions
Collaborative problem solving

Components

- Collaborative problem solving
  - Social skills
    - Participation
    - Perspective taking
  - Cognitive skills
    - Social regulation
    - Task regulation
    - Knowledge building
Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the “adequacy” and “appropriateness” of “inferences” and “actions” based on test scores or other modes of assessment.
ASSESSMENT

= Feedback for Teachers

= Teachers as Evaluators

= Making Learning Visible
Influences on Achievement

Decreased  Zero  Enhanced
What is the typical effect across?

- 900+ meta-analysis
- 50,000 studies, and
- 240+ million students
Effect on Achievement

Typical Effect Size

- Decreased
- Zero
- Enhanced

Size:

- 0
- 1.0

Effect Size: .40
Distribution of effects
<table>
<thead>
<tr>
<th>Rank</th>
<th>BOTTOM Influences</th>
<th>Studies</th>
<th>Effects</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>113</td>
<td>Class size</td>
<td>113</td>
<td>802</td>
<td>.21</td>
</tr>
<tr>
<td>114</td>
<td>Charter schools</td>
<td>18</td>
<td>18</td>
<td>.20</td>
</tr>
<tr>
<td>125</td>
<td>Matching learning styles</td>
<td>244</td>
<td>1234</td>
<td>.17</td>
</tr>
<tr>
<td>130</td>
<td>Mentoring</td>
<td>74</td>
<td>74</td>
<td>.15</td>
</tr>
<tr>
<td>131</td>
<td>Ability grouping</td>
<td>500</td>
<td>1369</td>
<td>.12</td>
</tr>
<tr>
<td>133</td>
<td>Gender</td>
<td>3168</td>
<td>6293</td>
<td>.12</td>
</tr>
<tr>
<td>134</td>
<td>Teacher education</td>
<td>106</td>
<td>412</td>
<td>.12</td>
</tr>
<tr>
<td>136</td>
<td>Teacher subject matter knowledge</td>
<td>92</td>
<td>424</td>
<td>.09</td>
</tr>
<tr>
<td>145</td>
<td>Class architecture</td>
<td>315</td>
<td>333</td>
<td>.01</td>
</tr>
<tr>
<td>148</td>
<td>Retention</td>
<td>229</td>
<td>2882</td>
<td>-.13</td>
</tr>
<tr>
<td>Rank</td>
<td>Influences</td>
<td>Studies</td>
<td>Effects</td>
<td>ES</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------------------</td>
<td>---------</td>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>41</td>
<td>Study skills</td>
<td>668</td>
<td>2217</td>
<td>.59</td>
</tr>
<tr>
<td>43</td>
<td>Outdoor/ Adventure Programs</td>
<td>187</td>
<td>429</td>
<td>.52</td>
</tr>
<tr>
<td>46</td>
<td>Play Programs</td>
<td>70</td>
<td>70</td>
<td>.50</td>
</tr>
<tr>
<td>47</td>
<td>Second/Third chance programs</td>
<td>52</td>
<td>1395</td>
<td>.50</td>
</tr>
<tr>
<td>52</td>
<td>Early Intervention</td>
<td>1704</td>
<td>9369</td>
<td>.47</td>
</tr>
<tr>
<td>54</td>
<td>Pre school programs</td>
<td>358</td>
<td>1822</td>
<td>.45</td>
</tr>
<tr>
<td>57</td>
<td>Teacher expectations</td>
<td>674</td>
<td>784</td>
<td>.43</td>
</tr>
<tr>
<td>83</td>
<td>Effects of testing</td>
<td>569</td>
<td>1749</td>
<td>.34</td>
</tr>
<tr>
<td>92</td>
<td>Test-based accountability systems</td>
<td>92</td>
<td>92</td>
<td>.31</td>
</tr>
<tr>
<td>Rank</td>
<td>TOP influences</td>
<td>Studies</td>
<td>Effects</td>
<td>ES</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>Student expectations</td>
<td>209</td>
<td>305</td>
<td>1.44</td>
</tr>
<tr>
<td>3</td>
<td>Response to intervention</td>
<td>13</td>
<td>107</td>
<td>1.07</td>
</tr>
<tr>
<td>4</td>
<td>Teacher credibility</td>
<td>51</td>
<td>51</td>
<td>.90</td>
</tr>
<tr>
<td>5</td>
<td>Providing formative evaluation</td>
<td>30</td>
<td>78</td>
<td>.90</td>
</tr>
<tr>
<td>7</td>
<td>Structured classroom discussion</td>
<td>42</td>
<td>42</td>
<td>.82</td>
</tr>
<tr>
<td>10</td>
<td>Feedback</td>
<td>1310</td>
<td>2086</td>
<td>.75</td>
</tr>
<tr>
<td>12</td>
<td>Teacher-student relationships</td>
<td>229</td>
<td>1450</td>
<td>.72</td>
</tr>
<tr>
<td>25</td>
<td>Not labeling students</td>
<td>79</td>
<td>79</td>
<td>.61</td>
</tr>
<tr>
<td>48</td>
<td>Goals (Difficult vs. Do your best)</td>
<td>895</td>
<td>1111</td>
<td>.50</td>
</tr>
<tr>
<td>32</td>
<td>Worked examples</td>
<td>62</td>
<td>151</td>
<td>.57</td>
</tr>
</tbody>
</table>
Visible Teaching – Visible Learning

When teachers **SEE** learning through the eyes of the student and when students **SEE** themselves as their own teachers.
The Power of Feedback

Average effect = .40
• Constructive, commentary, correction, conversation, compliance to task
• **Give** lots of feedback
• Tells me “Where to next”

• Receive little feedback
RCT and Experimental studies

- Competence/Rubric
- Comparative/Social
- Individual

Clarity and Challenge of Success Criteria

NO FEEDBACK

- Including effort
- With Praise

FEEDBACK

Prior achievement
FEEDBACK FOR TEACHERS

rather than more testing of students

MODEL THE INSTRUCTION WE VALUE

rather than teaching to the tests we have

PUT ASSESSMENT WITHIN INSTRUCTION

rather than instruction into assessment

EMPHASIS PROGRESS TO STANDARDS/SUCCESS CRITERI

rather than have students met the standards
Must emphasize interpretation and Where to next?
Must have consequences
Must tell teachers where and for whom they have had impact success
Must relate to what has been recently taught relative to success criteria
Must relate to curriculum fidelity --concepts NOT items
Must be individual and group rich
Must include Progress as well as Proficiency
Must be immediate & tailored to the “here and now”
Must include Diagnostic as well as Success reporting
Must be based on modern test theory
Must have multiple modes of delivery
Must balance surface and deep processing
Must emphasize test information – particularly formative interpretations
Creating a test

Create Custom Test Step 1: Test Details

Please specify the following settings for your test. Fields marked with an asterisk [*] are mandatory.

Test Details

- **Test Name:** *Week 5 math test*
- **Description:** This is a test advancements in closing gaps in reading skills recognised in Week 2
- **Test Duration:** *30 Mins*
- **Subject:** *Reading*

Quick Help

- Enter a name for your test (max 20 characters).
- Provide a brief description of your test.
- Enter in the duration of your test. This must be a whole number e.g., 15, 20, 30.
- Select the Subject of your test. You will be able to specify the difficulty and Curriculum Strands on the following screens.

---

Reading, Panui, Writing, Tuhituhi, Numeracy Pangarua

---

[Cancel] [Continue >]
Create Custom Test Step 3: Curriculum Strands

Use the sliders to select up to 2 Curriculum Strands for your test.
Step 2: Difficulty Level

Please select up to 3 adjacent difficulty levels for your test.

DIFFICULTY LEVEL

LEVEL 2  LEVEL 3  LEVEL 4  LEVEL 5  LEVEL 6

Most  Many  Some  Few  None
Test characteristic curve
Step 4: Delivery Method

Please specify the delivery method of your test.

DELIVERY METHOD*

- Paper and Pencil Administered
- Onscreen Administered
- Computer adaptive Onscreen

ONSCREEN OPTIONS

Test Duration: 45 mins
Time to Review Test: 2 mins
Time for Attitude Questions: 5 mins
Total Test Time: 19 mins

Closed Questions:
- SOME
- MANY
- MOST
- ALL

Estimate of marking time per student: 2 mins
Creating the Test

Creating Test

Your test is being created.

This process may take some time, however, you can continue using this site while this process is occurring.
Welcome Kurt

Create a Test
To create a test, click the button below.

View Reports
To view reports, click the button below.

Assign a Test
To assign a test, click the button below.
Individual Learning Pathways

Learning Pathways Report for Test: Reading U, C, SF

Group: All Test Candidates
Student: Davis Crispeness

Date Tested: 22 October 2003

---

**Correct**

**Strengths**
- Make inferences: (15, 22, 33)
- Knowledge of vocabulary: (11, 20, 24, 28, 33)
- Respond using understandings & information: (11, 25)
- Scan for information: (19, 25)
- Find, select, & retrieve information: (10, 26)
- Punctuation: (15, 24)
- Make links between aspects of text: (15)
- Make use of prior knowledge: (23)
- Identification and understanding of main ideas: (20)

**Achieved**
- Respond using understandings & information: (2, 6, 13, 21)
- Scan for information: (2, 21)
- Find, select, & retrieve information: (2, 21)
- Knowledge of vocabulary: (5)
- Knowledge of semantic, syntactic, & visual grapho-phonics cues: (6)
- Identification and understanding of main ideas: (13)
- Understand & organise or sequence material: (2)

**aRs Score**

---

**Incorrect**

**To Be Achieved**
- Make links between verbal & visual information: (4, 5, 16)
- Respond using understandings & information: (10, 18, 23, 26, 29)
- Knowledge of poetic & figurative language: (10)
- Knowledge of vocabulary: (5, 7, 10, 31)
- Use grammatically correct structures: (7)
- Knowledge of semantic, syntactic, & visual grapho-phonics cues: (7)
- Make use of prior knowledge: (26)
- Knowledge of publishing/next conventions (e.g., Index, Contents): (26)
- Make links between aspects of text: (27, 29, 32)

**Gaps**
- Respond using understandings & information: (1, 8, 9, 12, 15)
- Identification and understanding of main ideas: (1)
- Find, select, & retrieve information: (1, 3, 15, 17)
- Use grammatically correct structures: (8, 9)
- Knowledge of semantic, syntactic, & visual grapho-phonics cues: (8)
- Knowledge of vocabulary: (8, 9)
- Understand & organise or sequence material: (3)
- Make inferences: (12)
- Make links between verbal & visual information: (12)

---

This student
Level
Year 5 mean

<table>
<thead>
<tr>
<th></th>
<th>Surface</th>
<th>Deep</th>
<th>Understanding</th>
<th>Connections</th>
<th>Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>aRs</td>
<td>430</td>
<td>466</td>
<td>408</td>
<td>419</td>
<td>379</td>
</tr>
<tr>
<td>Year 5</td>
<td>2P</td>
<td>2A</td>
<td>2P</td>
<td>414</td>
<td>2P</td>
</tr>
<tr>
<td>mean</td>
<td>452</td>
<td>464</td>
<td>446</td>
<td>448</td>
<td>440</td>
</tr>
</tbody>
</table>
Group Learning Pathways Report for Subject: Reading

Group: All Test Candidates
Group Size: 22

Number of Tests: 2
Period Tested: October 2003 to October 2003

### Key
- **Gaps %**
- **To Be Achieved %**
- **Achieved %**
- **Strengths %**

### Finding Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Gaps %</th>
<th>To Be Achieved %</th>
<th>Achieved %</th>
<th>Strengths %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find, select, &amp; retrieve information (No. Students = 22 No. Items = 26)</td>
<td>16</td>
<td>46</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Skim/scan for information (No. Students = 22 No. Items = 13)</td>
<td>14</td>
<td>50</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>Use reference materials, e.g., dictionary, thesaurus, atlas (No. Students = 15 No. Items = 1)</td>
<td>33</td>
<td>13</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>

### Knowledge

<table>
<thead>
<tr>
<th>Category</th>
<th>Gaps %</th>
<th>To Be Achieved %</th>
<th>Achieved %</th>
<th>Strengths %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge of poetic &amp; figurative language (No. Students = 15 No. Items = 1)</td>
<td>13</td>
<td>60</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Knowledge of publishing/text conventions (e.g., Index, Contents) (No. Students = 22 No. Items = 2)</td>
<td>3</td>
<td>57</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>Knowledge of semantic, syntactic, &amp; visual grapho-phonetic cues (No. Students = 15 No. Items = 3)</td>
<td>24</td>
<td>35</td>
<td>34</td>
<td>7</td>
</tr>
<tr>
<td>Knowledge of vocabulary (No. Students = 22 No. Items = 19)</td>
<td>17</td>
<td>45</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

### Understanding

<table>
<thead>
<tr>
<th>Category</th>
<th>Gaps %</th>
<th>To Be Achieved %</th>
<th>Achieved %</th>
<th>Strengths %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistently read for meaning (No. Students = 22 No. Items = 15)</td>
<td>4</td>
<td>56</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>Identification and understanding of main ideas (No. Students = 22 No. Items = 7)</td>
<td>20</td>
<td>36</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>Respond using understandings &amp; information (No. Students = 22 No. Items = 23)</td>
<td>28</td>
<td>38</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>Understanding of detail to support main ideas (No. Students = 22 No. Items = 2)</td>
<td>7</td>
<td>42</td>
<td>16</td>
<td>35</td>
</tr>
</tbody>
</table>

### Connections

<table>
<thead>
<tr>
<th>Category</th>
<th>Gaps %</th>
<th>To Be Achieved %</th>
<th>Achieved %</th>
<th>Strengths %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compare similarities &amp; differences within &amp; between texts (No. Students = 11 No. Items = 5)</td>
<td>3</td>
<td>54</td>
<td>13</td>
<td>30</td>
</tr>
</tbody>
</table>
This report is designed to answer the question “Where are students relative to the targets of Curriculum Levels 2 to 6”?

This report enables teachers to monitor the effect of teaching and learning activities on student progress within levels.
Target setting and expectations

Target Setting: Set Targets for December 2008

Targets for Reethu Xavier (Year 7)

Teacher or student target

Polynomial regression target
### The what next report

**What Next Report for Test: Geometry Y6 T3 2007**

**Group:** All Test Candidates

<table>
<thead>
<tr>
<th>Skill Level</th>
<th>Number Knowledge</th>
<th>Number Operations</th>
<th>Algebra</th>
<th>Measurement</th>
<th>Geometric Knowledge</th>
<th>Geometric Operations</th>
<th>Probability</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Advanced</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>6 Proficient</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>6 Basic</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>5 Advanced</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>5 Proficient</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>5 Basic</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>4 Advanced</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>4 Proficient</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>4 Basic</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>3 Advanced</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>3 Proficient</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>3 Basic</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>2 Advanced</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>2 Proficient</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>2 Basic</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

**Mathematics**

**Level 3 Advanced: Geometric Operations**

- Define plane shapes, prisms, pyramids, cones, and spheres by their spatial features.
- Represent objects with drawings and models.

**Position and orientation**

- Create and use rectangular and rotation-based co-ordinate systems to specify locations and describe paths.

**Transformation**

- Describe the transformation (reflection, rotation, translation, or enlargement), that has mapped one object onto another.


Please select resources in response to the average achievement of your group of students.

---

**Diagnostic advancement leveraging appropriately targeted online resources**
Additions to reporting

- Video assessment
- Essay scoring
- Complexity/ Cognitive process scoring

**SURFACE**

One idea. Who painted *Guernica*?

Many ideas. Outline at least two compositional principles that Picasso used in *Guernica*.

**DEEP**

Relate ideas. Relate the theme of *Guernica* to a current event.

Extend Ideas. What do you consider Picasso was saying via his painting of *Guernica*?
Additions to reporting

- Video assessment
- Essay scoring
- Complexity scoring
- Reporting progress
Additions to reporting

• Video assessment
• Essay scoring
• Complexity scoring
• Reporting progress
• Value added
Additions to reporting

- Video assessment
- Essay scoring
- Complexity scoring
- Reporting progress
- Value added
- Other domains

- Year 0-4
- Adult literacy
- Linked to LMS, to parents, to curricula change
- Other tests – spelling, commercial, health
Additions to reporting

- Video assessment
- Essay scoring
- Complexity scoring
- Reporting progress
- Value added
- Other domains
- World wide testing
Additions to reporting

- Video assessment
- Essay scoring
- Complexity scoring
- Reporting progress
- Value added
- Other domains
- World wide testing
- Cognitive processes

- Pattern recognition
- Developing production rules
- Different levels of integration of knowledge
- Different degrees of procedural skills
- Speed of processing
- Misconceptions
- Reports about cognitive strengths and gaps
- Activate instructional components
Gaussian distribution of errors

1788

Lord & Novick CTT to IRT

1968

Need for new model

2012 ??

Spearman start of CTT

1905

Hambleton & Van der Linden
IRT to MIRT

1998

The Future of Assessment

TEST                        ITEMS                        REPORTING ENGINE
Visible learning
Adolescent reputations & risk
Unlocking formative assessment
Assessing teachers for professional certification
Intelligence & Intelligence Testing

Forthcoming
Hattie, J. & Anderman E (Eds.)
Handbook of student achievement

Yates, G. & Hattie, J.
Making Learning Visible

jhattie@unimelb.edu.au