Australian Council for Educational Research (ACER)

From the SelectedWorks of Dr Ross Hudson

June, 2014

# Student Science Performance On Process Items Compared To Knowledge Items.

Ross D Hudson, Dr

Available at: https://works.bepress.com/ross_hudson/13/

# STUDENT SCIENCE PERFORMANCE ON PROCESS ITEMS COMPARED TO KNOWLEDGE ITEMS.

**Ross D Hudson**
Australian Council for Educational Research, Australia
ross.hudson@acer.edu.au

**Abstract** Assessment items in science have traditionally focussed on science knowledge. However, since literacy has become more important in science teaching and as a result science assessments have introduced a larger range of item s that are able to assess science investigative skills and science understanding as well as science knowledge. This in no small part can be attributed to the influence assessments such as PISA and to a lesser extent programs such as the TIMSS. The shift towards investigative or science process items raises the question of how this change will affect student performance. This research explored a series of large scale international population tests (International Benchmark Tests; IBT), that used instruments containing a balance of both science process and knowledge items. The instruments were administered to nearly 22,000 students, both primary and secondary aged. Rasch DIF analysis and correlation analysis showed little difference between the performances on the two item types. The conclusion from this study suggests that regardless of the students' ability they will perform equally well on science items regardless of whether the items have a process or knowledge focus. This result though somewhat surprising suggests that students are not disadvantaged by the changing focus of different questions

**Key Words** Assessment. Knowledge, Process, Population tests, Science

## STUDENT SCIENCE PERFORMANCE ON PROCESS ITEMS COMPARED TO KNOWLEDGE ITEMS.

## 1. Introduction

In recent times there has been a trend towards the use of assessment items in science that focus more on performance in science process (investigative) skills rather than performance on knowledge items. This in no small part can be attributed to the influence of the Program for International Student Assessment (PISA). This program has highlighted country differences in performance in the area of science process and has acted as a driver for governments in revising education programs and curriculum focus. Other programs such as the Trends in International Mathematics and Science Study (TIMSS) have concentrated more on assessing student understanding on predominantly knowledge items. However, country performances in these two programs do correlate well (Brown, Micklewright, Schnepf, & Waldmann, 2007; Gronmo & Olsen, 2008; Wu, 2009). This research explored a number of population tests, both national and international, that used instruments containing a balance of both science process and knowledge items.

The International benchmark tests (IBT) are a series of tests that are administered to a number of countries across the Indian ocean region including India, Indonesia, the Philippines, South Africa, UAE, Oman, Bahrain, Qatar, Kuwait, Saudi Arabia, Jordan.

These tests are developed by expert item writing teams from the Australian Council for Educational Research (ACER). The tests were benchmarked against the TIMSS scale and are historically linked from year to year allowing participating schools to measure progress within their school and at the same time measure performance against an internationally recognised scale (see http://www.acer.edu.au/tests/ibt). This situation is changing an an internal scale scale based on historical performances is being developed.

Items in these tests are written to match a generalised science curriculum appropriate to the TIMSS curriculum. In particular items are also designed to test both scientific knowledge and scientific investigation skills. These two concept areas match favourably with the other major international science test the PISA tests where questions are broadly classified as either Knowledge about science (investigative or conceptual skills) and knowledge of science(content understanding).
Comparisons between TIMSS and PISA have demonstrated that in spite of the different emphasis in balance between these two areas performance on the two tests appear to correlate well. (Brown, Micklewright, Schnepf, & Waldmann, 2007; Gronmo & Olsen, 2008; Wu, 2009). Specifically TIMSS has a larger focus on Knowledge of science compared to PISA. Furthermore analysis of the teaching methods within participating countries in PISA have shown that the top performing countries tend have a greater emphasis on practical skill development compared to lesser performing countries(Abell & Lederman, 2007)).
The PISA report for 2007 shows clearly that there is little difference in performance between students on the different science competencies (OECD (2009)

The IBT instruments have developed a balance between the two basic item types. The instruments each consist of 4 MCQ with usually around 40 % of items being classified by the writing team as being investigative in nature and the remaining 60 % being essentially knowledge based. Of interest was the relative performance on the two item classifications. The countries involved in the IBT assessments have tended to have fairly content driven curricula and teaching methods(Abell & Lederman, 2007)so it may have been expected that students may not perform as well on investigative items as they did on knowledge based items.

## Methodology

The IBT instruments were delivered to both primary and secondary aged students. Data from several years of 2009-2011 were obtained. The number of students participating in the assessments was 21 263. Alpha Reliability measures of the tests were calculated to demonstrate the validity of the items and the assessments over the period being considered. For the last year (2011) performance on all items was measured using Item response theory (IRT). Rasch analysis was used to find the item difficulty location for each item and also the mean student ability. Items were then separated and compared on the basis of whether they had been classified as I or K. The item locations of the I and K items were compared and measured against the IBT scale and more importantly the mean student location (performance) on the two item classifications compared.

## Results

The reliability of the science tests ranged from 0.78 to 0.84.
Generally, in terms of item difficulty the Investigative items were little different than the knowledge items. Students found the items equally difficult or easy. None of the differences in mean item difficulties were not significant. However, the range of item difficulties was greater in all grade levels for the knowledge items as compared to the investigative items. That is the knowledge items proved be the most difficult and also the easiest items. Figure 1 shows that whilst the mean difficulties of the investigative and knowledge items types were virtually same the range of the knowledge items was greater. Table 1 shows the summary statistics for the grade 6 item types.

**Figure 1. IBT Science Grade 6 - K and I item types**

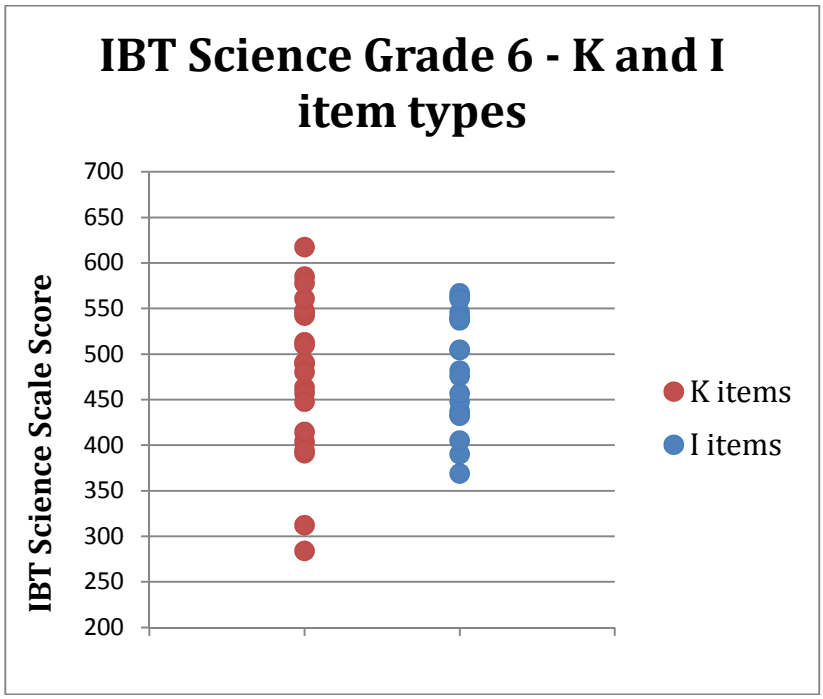


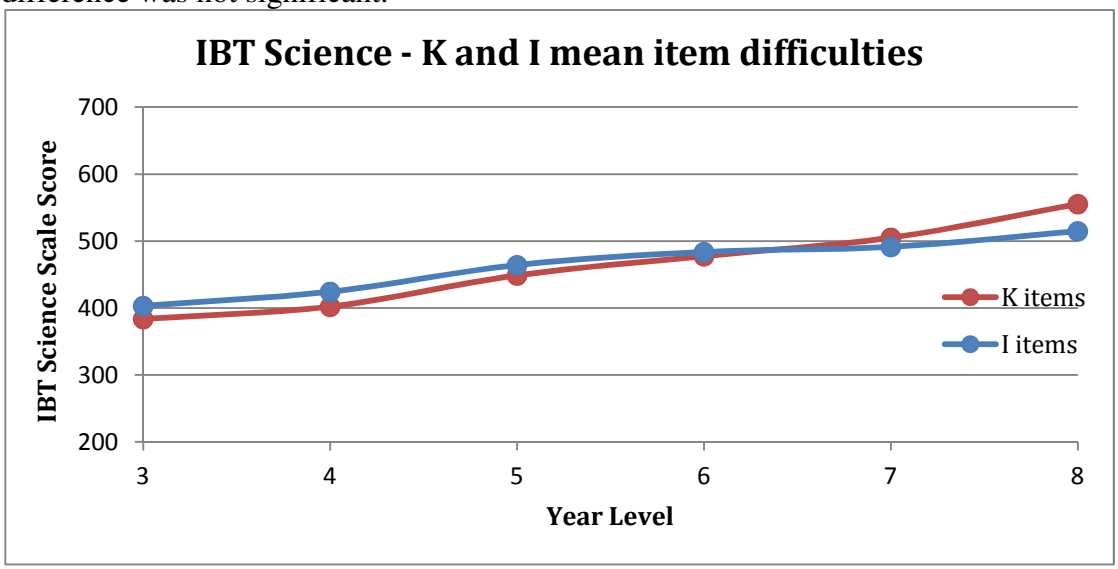

**Table 1.**. IBT Science Grade 6 - K and I item difficulties

| IBT Science Grade 6 - K and I item difficulties | | | |
|---|---|---|---|
| | N | Mean | Std. Dev. |
| K items | 21 | 477.6 | 87.9 |
| I items | 19 | 483.8 | 62.5 |
| | | T-test p-value | 0.797 |

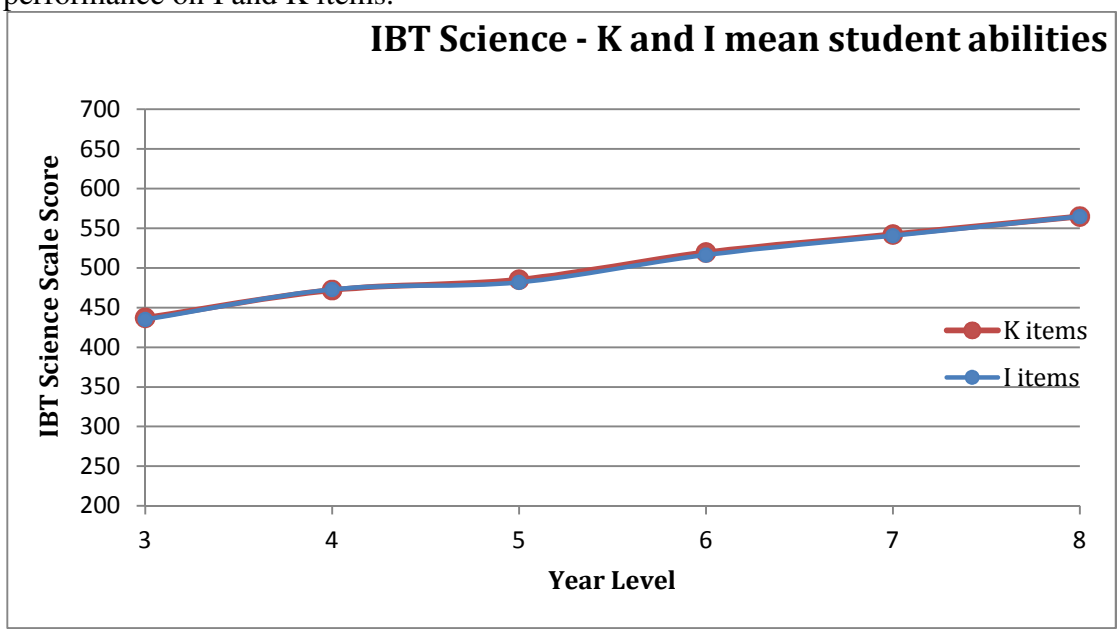
| IBT Science Grade 7 - K and I item descriptives | | | |
|---|---|---|---|
| | N | Mean | Std. Dev. |
| K items | 23 | 505.0 | 62.4 |

| | | | |
|---|---|---|---|
| I items | 17 | 491.4 | 45.5 |
| | T-test | p-value | 0.429 |

When comparing student performance on the two item types there was again little difference in performance. That is students performed equally well on either item type. Again the difference was not significant.




The overall result showed that investigative items at the lower levels (grades 3,4 & 5) investigative items proved slightly more difficult . At Year 6 & 7 there was little difference and by Year 8 the knowledge items were tending to a higher degree of difficulty.
In terms of student abilities there was virtually no difference in student abilities as in terms of performance on I and K items.




In summary there was virtually no difference in student performance on investigative compared to knowledge items. Furthermore, despite a wider difficulty range with knowledge items there was no appreciable difference between the relative difficulties of knowledge items compared to investigative items.

## Conclusions

The IBT tests mirror some aspects of both the PISA tests and the TIMSS tests. The IBT tests are benchmarked against the TIMSS so schools and students can easily see how they compare to the TIMSS performance of different countries. Whilst a direct comparison between IBT and PISA is not yet possible it is interesting to note that the performance characteristics of the two assessments in terms of performance on items that examine different science competencies is similar. The research shows that students perform equally well on either investigative and knowledge based questions .The measured difficulty location range on knowledge items was greater than that of the investigative items, but these differences were not significant. This suggests that the IBT assessments provide a valid measure of student abilities in both the investigative competency and the knowledge competency.

## References


Abell Sandra K., Lederman Norman G.(2007) *Handbook of Research on Science Education* (1st. ed.). Mahwah, N.J.: Lawrence Erlbaum Associates,

Brown, G., Micklewright, J., Schnepf, S., & Waldmann, R. (2007). International surveys of educational achievement: how robust are the findings. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(3), 623-646.

Gronmo, L., & Olsen, R. (2008). *TIMSS versus PISA: the case of pure and applied mathematics*. Retrieved February, 2008, from http://www.timss.no/publications/IRC2006_Gronmo&Olsen.pdf

OECD (2009),*Equally prepared for life?: How 15-year-old boys and girls perform in school*, PISA, OECD Publishing. doi: 10.1787/9789264064072-en

Wu, M.L. (2008). *A Comparison of PISA and TIMSS 2003 achievement results in Mathematics and Science*. Paper presented at the Third IEA Research Conference, Taipei, September, 2008.

The proposal introduces a paper or a poster presentation, aimed to present in the NFSUN conference and it should be written according to the APA (American Psychological Association) Publication Manual format and include: title, author(s), the institute and e-mail address of the corresponding author; an abstract; text; and references. The proposal should be 2 to 3 pages large and not more than 1,600 words. The proposal should be written in English.

This proposal template should be downloaded, prepared according to the guidelines found in this document and then submitted through a submission form https://elomake.helsinki.fi/lomakkeet/42679/lomake.html before January 31st, 2014.

Please, name your proposal document as: **Lastname_Firstname_Proposal_Nordic_2014**. The proposals should be submitted as Microsoft Word file (.doc or .docx) or OpenDocument Text file (.odt). No other file format will be accepted.

Proposals will be reviewed by two referees. Names of the authors will be reduced before the proposal is submitted to the referees. The paper and poster presentations to the conference will be selected based on the review. The referees could suggest a paper presentation to be changed as a poster presentation. The review process requires some weeks and it will be completed by March 10th, 2014. After that, the decision will be sent to the authors of the proposals and the conference program will be formed.

The accepted proposals could be enlarged after the conference and submitted as papers to the conference proceedings (a special issue of LUMAT journal http://www.luma.fi/lumat-en). Detailed information about the submission of full papers is given in the conference.

Paper size should be A4 (210 mm x 297 mm) and all margins should be as in this document. The following guidelines should be followed in the formatting of the text: Font: Times New Roman, size 12, spacing 1, empty line between paragraphs.

Headings are helpful as a guide for readers. Three or four levels of headings are usually sufficient. Please, use heading styles of your word processing software.

In your paper, use a first-level heading only for the main title. Use second-level headings for the title which appears at the beginning of your introduction and for the words Method, Results, Discussion and References which appear at the beginning of the appropriate sections.

Use a third-level heading for each subsection of your paper (e.g., subsections of your introduction).

Number headline levels from second level as 1., 1.1., 1.2., 2.…

Examples:

# Document Title (Arial 18 pts, Heading 1 style)

## 1. Second-Level Heading (Arial 14 pts, Heading 2 style)

### 1.1. Third-Level Heading (Arial 11 pts, Heading 3 style)

#### *1.1.1. Fourth-Level Heading (Times New Roman 12 pts, Heading 4 style)*

Charts can be included in the document. Images should be submitted via webform as separate files and their places marked in the manuscript (e.g. <<add here image1.jpg>>)

*If the paper presents empirical research*, it should contain down into four primary sections: 1. Introduction; 2. Method; 3. Results; and 4. Discussion and Conclusions. Particular, it must be clear what the results of the study are and how these emerge from the analysis. Tables, graphs etc. must be clear and easy to read.

*If the paper presents theoretical research*, the argumentation must be clear and easy to follow, and conclusions must follow clearly from the arguments that are presented. Although this guide is intended for empirical papers, many of the ideas can also be followed in theoretical ones. It is especially important to present a clear formulation of the purpose of the paper, in which the research aims are stated and the study is placed in context with the broader field of research, with references to relevant research literature. Discussion, conclusions and/or implications must be clearly connected with the aims and the results (and/or the argumentation) presented in the study. Results must be discussed in light of relevant literature.

## 1.1 How to write the introduction

The introduction should present the topic of the paper, review relevant theories and research and brief provide a review of the research design that will follow, including research questions or hypotheses. It is not perhaps useful to use the title "Introduction" at the beginning, use your “own” title as a second level heading to start this section. Typically an introduction has three primary components.

The first component is the problem statement: describe the topic of interest and state briefly why investigating this topic is important (i.e., its significance for theory or practice).

The second component is the literature review: review some relevant theories (can be found in textbooks, handbooks and international journals) and/or earlier research findings that relate to your topic. Describe techniques that have resulted from the theoretical background provided. When you present research provide a brief description of its intent, the method employed, results and conclusion. Make certain you cite properly (see below).

The third component of the introduction is a statement of intentions. Briefly describe your experiment, relating it to your literature review. Briefly state your research questions or hypotheses.

## 2. Method

The Method section should describe the procedure in detail sufficient to allow those who wish to replicate the method to do so. This section should include a description of the participants (teaching experiment(s) ...), and the research methods used. Report the number of participants, age, gender and important demographics that relate to the experiment. Describe the research method (procedure) so that a reader could precisely replicate the study or experiment. You may choose to put a copy of the questionnaire or interview protocol in an appendix. Illustrations of apparatus or settings should be presented as figures (see on next chapter for the presentation of figures). Each figure should be referred to using a number. When referring to a figure within the text say something such as, "Figure 1 presents..." or "The pupils worked in pairs and ... (see Figure 1)".

## 3. Results

The “Results” section of your paper present (empirical) results without interpretation. The only thing that should be described are the results, including descriptions of graphs of data obtained. Interpretation of your results and opinions on how they relate to your research questions or hypotheses should be placed in the discussion section.

Figures may be pictures, charts or schemes but are most commonly graphs of data. If the figure is a graph, each axis should be properly labelled. Make sure that the scales of the x- and y-axes represent the highest and lowest possible scores. Use a minimum of Times New Roman 12 pts in figures. All figures must be computer-designed and submitted as embedded pictures in your document (see in yours Help-menu how to “Insert picture” in “MS Word format”). It is recommended that digitalised photographs have 256 level greyscale.

Tables should be made as described in the examples below. Use a minimum of Times New Roman 12 pts in tables.

Figure and table captions should be short summaries of what is presented in each figure or table. Use the word "Figure" or “Table” and its associated number followed by a full stop (period). On the same line add a short description of the figure or table. For example, it is typical to introduce in the graph-labels the x and y axes (without actually using the terms x and y axes) and the legend (if there is one). Each figure should be referred to using a number. When referring a figure or a table within the text use expression such as, "Figure 1 presents..."

Example of tables:

**Table 1.** Evaluation of the learning materials teachers used to teach the basics of electronics and electricity.

| Evaluation of learning materials | Study group ($n$ = 81) | Control group ($n$ = 53) |
|---|---|---|

| | frequency | relative frequency | frequency | relative frequency |
|---|---|---|---|---|
| Poor | 14 | 17.3% | 8 | 15.1% |
| reasonably good | 17 | 21.0% | 11 | 20.8% |
| Good | 41 | 50.6% | 13 | 24.5% |
| no answer | 9 | 11.1% | 21 | 39.6% |

$\chi^2 = 17.1$***

Example of figures:

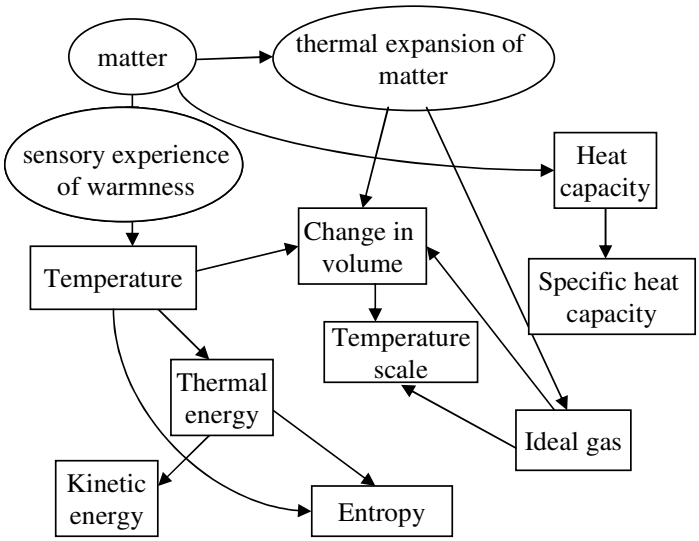



**Figure 2.** Example of a diagnostic network representation supplied by an individual student.

State the statistical analysis being used and what is being compared. For example, tell whether you are using ANOVA, state the means of the items being compared, then state your statistical conclusion in terms of statistical significance and report the statistical information at the end of this concluding sentence. For example: "There was a significant difference between the groups' responses (F(28)=3.12, p<.05)". Once again, do not comment on the results, but just describe what happened. Save your interpretation for the Discussion section.

*Appendix*: Label each component of the appendix using a letter designation (e.g., Appendix A). Refer to the appendix in the text in parentheses at the end of a sentence. For example: (see Appendix A).

## 4. Discussion and conclusions

The Discussion section is where you provide your interpretation of the results. Was the hypothesis (or hypotheses) supported or not? Quite often a paper finishes with a proposal future work. Overall, the architecture of a research paper can be viewed as a vase. The introduction starts broad and provides a stable foundation for the information to come, the paper then narrows to its main points and finally broadens again, fanning out into new horizons. It is therefore important to state potential future research questions and design at the end of the paper.

## 5. Acknowledgement

A brief acknowledgement section may be included before the reference section.

## 6. References

Make sure that all the citations in your paper are correctly referenced in the Reference section. The Reference section is placed at the end of the manuscript and is headed by the term "References". Entries should be listed in alphabetical order. Second and all following lines should be indented.

*Examples* of the most common types of reference book, articles, and chapters from an edited book can be found below. Be sure to pay attention to punctuation, capitalization, special formatting such as italics and other minor details (for example, only supply the initials of each author & not his/her first or middle name). Look carefully at the contents and the example to be sure you understand how to format each reference.

***Books***
*General Contents*

Author's name(s). (Publication date). *Name of Book* (Edition number if appropriate). Location of publishing Company: Name of publisher.

*Example*

Goleman, D. (1987). *The meditative mind: The varieties of meditative experience*. New York: St. Martins.

***Journal Articles***
*General Contents*

Author's name(s) (Publication date). Title of article. Name of journal, Volume number, page numbers.

*Examples*

Lukoff, D., Lu, F., Turner, R., & Gackenbach, J. (1995). Transpersonal psychology research review: Researching religious and spiritual problems on the Internet. *Journal of Transpersonal Psychology*, *27*(3), 153-170.

Maslow, A. (1962). Lessons from the peak experiences. *Journal of Humanistic Psychology*, *2*(3), 9-18.

***Book Chapter***
*General Contents*

Author's name(s) (Publication date). Title of chapter. In editor's name(s) (Eds.), *Name of book* (page numbers). Location of publishing Company: Name of publisher.

*Example*

Wilber, K. (1980). Eye to eye: Science and transpersonal psychology. In R. N. Walsh & F. Vaughan (Eds.), *Beyond ego* (pp. 62-100). Los Angeles: J. P. Tarcher.