acara AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

# Collaborative problem-solving online assessment and the Australian Curriculum

## Final report

**February 2018**

NSW GOVERNMENT Education Public Schools

Australian CURRICULUM

# Contents

Background ..... 3
The project ..... 3
Trialing ..... 4
Findings and analysis ..... 4
Issues identified from analysis ..... 5
Conclusion and future directions ..... 6
Appendices ..... 7
Appendix A: Observations from the trial and comments from the Cognitive interview process – excerpt from the Interim report (2016) ..... 7
Appendix B: Preliminary data analysis ..... 10
Appendix C: The marking/coding scheme and the data collected in detail ..... 24
Appendix D ..... 35

# BACKGROUND

The Collaborative problem-solving online assessment and the Australian Curriculum project aimed to develop innovative high-quality assessment that would elicit valid, reliable data on students' collaborative problem-solving skills. The Australian Curriculum, Assessment and Reporting Authority (ACARA) and the New South Wales Department of Education (NSW DoE) have worked in partnership with Janison (a learning and assessment platform provider) to develop new forms of assessment for general capabilities. The partnership has recognised that this is a shared challenge, and a better solution will be achieved through collaboration.

## The project

The project has achieved a high level of success in sharing knowledge and expertise in online assessment of collaborative problem-solving skills, particularly with educational partners, the Swedish Association of Local Authorities and Regions (SALAR), the National Council for Curriculum and Assessment (NCAA) Ireland, and the technical partner Janison.

A major component of the project was the development of prototype online tasks to assess collaborative problem-solving using an online management system that allows automatic collection of data while students interact with a randomly selected partner via a chat box. The tasks developed for the assessment were dynamic and interactive in nature and were presented to students in a rich real-world context via a virtual scenario.

The writing and online production of the tasks, together with the creation of the Australian Curriculum Collaborative Problem-solving Assessment Framework (ACCPSAF) and the marking/scoring rubrics for each task, have enriched the partners' understanding of how collaborative problem-solving may be addressed in online assessments.

The elements of the Framework were informed by the Australian Curriculum: General Capabilities, criteria from the Assessment and Teaching of 21st century skills (ATC21S), Hesse et al. (2012) and the PISA 2015 draft Collaborative Problem-solving Framework.

The criteria for the framework have been developed from a number of sources, including Krathwohl's Taxonomy of Affective Domain, *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy, Phases of Problem Solving,* Mayer and Wittrock 1996 *and The Structure of the Observed Learning Outcome Taxonomy.*

A unique aspect of the project was the simultaneous capture of students' actions and text responses while they were interacting with a partner, sometimes located in different schools and communities. The ability to capture students' thinking and communication with others, along with indicators of their level of knowledge and skills, have the power to transform models of assessment.

## Trialing

A proof of concept trial of one task was conducted in July 2015 with 900 Year 9 students in nine NSW schools, where students were experienced in online testing. The trial schools were selected to provide representative samples of students.

In October 2015, a full trial of four tasks was conducted in 20 schools (across urban, regional and remote locations) from five Australian states and territories with over 2,200 Year 8 students (approximately 14 years old). Data captured from the trial provided the evidence to locate students on the levels proposed in the assessment rubric. The trial included a cognitive interview process with students to provide developers with additional information about the design of the tasks and particularly about the marking/coding scheme for items. The cognitive interview process also provided information about the pairing procedures and the success of these approaches.

### Findings and analysis

1. Observations were recorded at all trial sessions in the following areas:
   - student engagement and teacher/school support
   - technical issues
   - timing
   - random pairing
   - cognitive interview process.

   Observations from the trial are detailed in **appendix A.**

2. A summary report from the cognitive interview process is also included in **appendix A**.

3. Student survey report

   The data from the student survey report have not been received. Based on the data gathered in the 2015 full trial a more complete analysis of this survey data would make it possible to compare the responses of different groups of students. This would provide valuable information about whether, for instance, students who felt that they had worked through problems in an organised way were successful in solving the problems in the tasks, or whether students who felt that they had suggested ideas themselves and used the ideas of others were rated highly in the social domain. This type of information would be very useful to teachers and students in providing feedback on how to solve problems collaboratively.

4. Data analysis

   The project has collected large and complex data files, capturing individual students' actions and responses. The data analysis was provided by the New South Wales Department of Education. Whilst the analysis is not a formal evaluation of the efficacy of the framework, it provides insights into how the tasks could be further refined.

   Several models have been applied to the data to determine the most effective method to assess student levels of achievement in collaborative problem-solving.

The analysis included treating criteria as:

- individual dichotomous items
- polytomous items (sum of scores for indicator)
- polytomous items (maximum score for indicator).

The full analysis including summary tables and figures, table of contents of additional item level summary statistics provided in Excel, and table of contents of additional support files (for example, code data, syntax) is provided in **appendix B.**

## Issues identified from analysis

The data were initially supplied in a long file form. There was a record for each task part for all student pairings. The data were restructured to create a single record per student pairing, with variables for that pairing for each task and associated relevant criteria.

A few students dropped out due to technical issues, therefore some students participated in multiple student pairings. As a result of these multiple pairings, multiple attempts at some of the tasks were recorded. In these preliminary data, no adjustments were made for the multiple student pairings and all student pairings were retained.

Analysis could detect a difference between non-attempts and zero scores, but not between items a student did not attempt or did not complete due to lack of time or technical issues.

A significant number of problems were identified in relation to coding. For instance, coding of a number of criteria, particularly in the task ‘Lights out, part A’ and in the ‘Windmill’ task, did not appear to reflect the scoring rubrics provided. The problems with coding appeared to occur when both actions and communications were recorded. For some criteria, all responses were coded as 0. It would be important to investigate whether certain actions could not or were not recorded by the system, for example, whether those criteria were coded as 0 (all incorrect), or some were not attempted.

Problems have been identified where the data (item characteristic curves) did not support the suggested level of difficulty of the action or communication described in the Framework. The scoring rubrics were developed directly from the Framework.

It should be noted that the task ‘How to vote’ had the highest number of increasing item characteristic curves and yet was considered by the test developers to be the least effective task in involving students in problem-solving or in collaborating with another student.

A detailed study of one of the tasks was conducted to better understand the issues arising from the analysis. This highlighted two of the issues mentioned above, including where criteria were coded as 0, and where the level of difficulty, as determined by the data analysis, did not match the level suggested in the scoring/marking scheme. This detailed study of ‘Lights out’ is provided in **appendix C**.

Observations from the National Council for Curriculum and Assessment (NCAA) Ireland can be found in **appendix D**.

## Conclusion and future directions

The purpose of this project was to build on the work of the Assessment and Teaching of 21st Century Skills Project (ATC21S) and PISA 2015 Collaborative Problem-solving (CPS) assessment, to develop an assessment framework aligned to the Australian Curriculum general capabilities (aspects of Critical and Creative Thinking, and Personal and Social Capability) and to identify further areas for development and research.

A critical aspect of the project was the desire to assess collaboration and problem-solving between students completing tasks in more authentic situations. The need for further research to achieve this goal is evident, in particular in terms of selecting working pairs, the effect of random pairing versus selected pairing, as well as the impact of age and maturity on engagement and collaboration.

The analysis of the data collected from the October 2015 full trial of four tasks has provided information that will inform how tasks of this kind might be further developed. However, based on the available data set and data analysis, and without deeper research, it is difficult to:

- validate the Framework and support future verification of aspects of the Australian Curriculum general capabilities, including their strength in describing collaborative problem-solving
- estimate the effect on the data (and therefore the validity of the analysis) of the number of students not completing the tasks and the number of students who repeated parts of the task with different partners
- provide definitive statements about the Framework due to the problems identified regarding the coding of some criteria. Research into the mechanisms used to capture and score actions and communications, and any problems experienced in the recording of this process would be considered an essential starting point for any future work.

The analysis does not allow a formal evaluation of the efficacy of the Framework. However, the data do suggest the criteria and the levels in the Framework related to the Social domain would benefit from further research into:

- how students communicate online and particularly how they communicate in the process of problem-solving
- the impact of partner status in students' problem-solving communication would provide important insights into the selection of appropriate criteria.

The levels in the Framework related to the Cognitive domain would benefit from further research into the strategies students use to problem-solve, particularly the processes used by students working with digital representations online.

# Appendices

## Appendix A: Observations from the trial and comments from the Cognitive interview process – excerpt from the Interim report (2016)

### Trialling and preliminary findings

A proof of concept trial of one task was conducted in July 2015 with 900 Year 9 students in nine NSW schools, where students are experienced in online testing using a similar environment. The trial schools were selected to provide representative samples of students.

In October 2015 a full trial of four tasks was conducted in 20 schools (across urban, regional and remote locations) from five Australian states and territories with over 2,200 Year 8 students (approximately 14 years old). Data captured from the trial provide the evidence to locate students on the levels proposed in the assessment rubric. The trial included a cognitive interview process with students to provide developers with additional information about the design of the tasks and particularly about the marking/coding scheme for items. The cognitive interview process also provided information about the pairing processes and its success.

### Observations from the trial

Observations were recorded at all trial sessions in the following areas: student engagement and teacher/school support, technical issues (in the school and in the online environment), pairing and timing.

### Student engagement and teacher/school support

Based on anecdotal evidence provided by teachers, observers and survey data, the vast majority of the students who took part in the trial were fully engaged in the process and enjoyed the tasks. The assessment was undertaken by Year 8 students; however, there may be benefit in trialing with younger and older students. Staff were generally very supportive and helpful during the trial process. Some schools were eager to have their students participate and to find out how their students responded to the challenge.

### Technical issues

About half of the schools experienced some technical problems related to the computers that were being used or the system used to access the internet. For example, where the students brought their own devices, these were not always well-maintained with up-to-date software and settings. Although a number of schools used Wi-Fi to access the internet, this worked efficiently in some schools but not in others. With the initial schools during the trial, some issues were experienced with the online environment. Some of these issues were rectified later in the trial.

## Timing

For most of the schools involved, small numbers of students were able to complete all tasks in a reasonable time. Issues that contributed significantly to this overrun of the timing related to available computers and to students dropping out of a pairing situation and having to log back in.

## Random pairing

In the PISA 2015 draft collaborative problem-solving tasks, students collaborated with an avatar (computer agent) using pre-programmed responses and not with another student. The advantage of using an avatar with pre-programmed responses is that it makes for much easier/simpler automation of scoring and analysis; however, it is labour-intensive for developers.

To make the tasks more economically sustainable and to avoid the complexity of selecting working pairs, the project partners were particularly interested in the viability of students being paired randomly across schools via the online environment. It was presumed there was a symmetry of status, knowledge and roles with a shared problem space.

Students were paired randomly and across schools where possible. It was technically difficult to ensure students were paired across schools and to maintain consistent pairing as students were easily able to drop out of their pairing situations. There were a small number of selected pairings as part of the cognitive interview process.

The selection process for working pairs may be varied in future trialing to investigate the effect of random pairing versus selected pairing.

## Cognitive interview process

The cognitive interviews in this research activity focused on the cognitive and social skills that students used to solve the problems presented in the tasks. Information was collected on students' interaction and engagement with the four assessment tasks and in particular students' feedback about the interaction between the two students. Interviews examined the extent to which the online tasks assess knowledge and skills not easily assessed by more common testing programs. The cognitive interviews were conducted by an external consultant, Education Assessment Australia (EAA).

EAA reported that:

> Though all students interviewed used computers regularly both at home and school, the level of technological sophistication in the assessment came as a pleasant surprise to most… Their satisfaction with the collaborative aspect of the assessment was most pronounced where they had a partner who was responsive and able, and being paired with a partner who was less able or disengaged could be "*frustrating*"…. Chat-based communication with a partner in the form of typing was not problematic… Though not specifically designed as a measure of a student's capacity for effective micro learning, the game- or puzzle-like nature of the tasks was frequently observed to involve significant amounts of micro learning in order to complete them. (EAA, 2015: 12)

Micro-learning referred to in the EAA report can be considered as part of the problem-solving process described as barriers by Frensch and Funke (1995). They describe the ‘givens’ as the knowledge that the student has about the problem at the outset and ‘barriers’ as lack of knowledge, process or strategies that stand in the way of achieving the solution.

The barriers must be overcome using the available tools in the assessment. The data may provide further insights into the impact of micro learning on collaborative problem-solving.

The findings from the cognitive interviews will assist in validating the four tasks and the Australian Curriculum Collaborative Problem-solving Assessment Framework.

## References

Educational Assessment Australia (EAA) 2015, *Collaborative Problem Solving Online Assessment 2015 Trial: Cognitive Interviews Report*, unpublished.

Frensch, PA and Funke, J (eds) 1995, *Complex Problem Solving. The European Perspective,* New York: Psychology Press.

# Appendix B: Preliminary data analysis

Prepared by Analytics, High Performance, NSW Department of Education

## Overview

### Primary purpose

To provide a brief ‘first look’ and commentary on the tasks and the potential of these types of collaborative assessments. The goal is not to try to establish a formal scale, but to give some information towards how you might further develop tasks of this kind.

### Secondary purpose

Examine dichotomous, maximum set score and sum score approaches to scoring. This approach broadly follows the initial proposed assessment framework but is not a formal evaluation of the efficacy of the framework or ways to improve it.

### Outline of the current report

**Tab A: Summary of method and results**

Brief description of initial approaches to scoring and summary of preliminary analysis in words.

**Tab B: Summary tables and figures**

**Tab C: Table of contents of additional item level summary statistics provided in Excel**

**Tab D: Table of contents of additional support files provided (e.g. code data, syntax)**

# Tab A: Summary of method and results

### Overview of the tasks

The CAA pilot in 2015 contained four tasks:

- Lights Out
- Let's Tackle the Fish
- Who's Got Your Vote
- Windmill

For a breakdown of elements, indicators and criteria assessed in each task, please see Table 1.1 in **Tab A**.

## Overview of preliminary analysis methodology

### Structure of the data

The data was initially supplied in long file form. There was a record for each task part for all student pairings. The data was restructured to create a single record per student pairing with variables for that pairing for each task and associated relevant criteria.

As a result of issues with students dropping out, some students participated in multiple student pairings. In this preliminary data, no adjustment was made for the multiple student pairings and all student pairings were left as-is.

### Treating criteria as individual dichotomous items

Criteria not applicable to a task were not included in the final item set for Rasch analysis.

Extreme criteria (with all valid responses incorrect) were not included in the Rasch analysis, however facility rates are provided. See Table 1.2 for a list of removed criteria.

For example:

- For the three criteria assessed for SOM (Social Management) "a" (communicate effectively) in Lights Out Part A, all valid responses were coded as 0 for the three applicable criterion. This suggests that students did not demonstrate at least one of the following:
    - Something typed into the chat box, but not using suggested text
    - Responding with an appropriate greeting using "Hello", "Hi", "My name" or "Hey"
    - Generating communication by using "How you", "who you", "where you", "what you"
    - A coding error?
- For the CI-a1 criterion – "Cognitive" strand, "Inquiring" element, "Collects and organises information" indicator in Windmill Part B, all valid responses were coded as 0. This suggests that students did not demonstrate:
    - Random actions undertaken/did something
    - A coding error?

Three indicators did not receive correct responses for any of their criteria for a particular task and were removed from all treatment methods:

- The Social management (SOM) a (communicate effectively) criteria for Lights Out Part A
- The Self-management (SEM) a (Uses self-discipline and sets goals) criteria for Lights Out Part A
- SOM criteria a (communicate effectively) for Windmill

Within the criteria that had all valid responses coded as incorrect, 69% were an action criterion (required a click or action) and 31% were criteria that involved the chat box (requiring text typed). The ratio of action versus chat criteria was similar for items with a mixture of correct and incorrect responses (i.e. able to be entered into Rasch analysis).

Item Characteristics Curves were generated for each criterion relevant to a task that did not receive only “incorrect” responses. These criteria were classified into eight categories based on the appearance of the ICC (see Figures 1.1 to 1.8 for exemplars). For a summary of the ICCs of each task see Tables 1.3 to 1.6, and for the assessment overall see Table 1.7. Summaries of the ICCs for each criterion and for each indicator can be found in the Excel workbook (See **Tab C** Excel workbook outline).

- Over 50% of the criteria within Lights Out v6 had non-increasing ICCs, or were all incorrect.
- 44% of the criteria within Let’s Tackle the Fish had non-increasing ICCs, or were all incorrect.
- Only 15% of the criteria within Who’s Got Your Vote? had non-increasing ICCs. However, 55% of the criteria had ICCs that showed higher levels of discrimination compared to the rest of the assessment. Who’s Got Your Vote? also had the highest percentage of ICCs that were increasing, at 85%.
- Over 60% of the items within Windmill had non-increasing ICCs, or were all incorrect.

Item difficulty ranged from -9 to 9 logits. Most items with an item logit less than -5 showed item characteristics curves with probabilities of 1 or almost 1 for all class intervals. Most items with an item logit more than 3.8 showed item characteristics curves with probabilities of 0 or almost 0 for all class intervals.

The person-Item location map for individual criteria is presented in Figure 1.9

Item characteristics curves for all criteria are presented in **Tab D**.

**Possible recommendation:**

Since students did not achieve any of the criteria within the three indicators above, this suggests that the coding of these criteria needs further investigation, this may be a result of response coding, or the framework itself. Further investigation may also be necessary for other criteria receiving no correct responses.

Of the four tasks, Who’s Got your Vote did not require any item deletions and had the highest number of increasing item characteristics curves, suggesting that this task was most successful when treating each criterion as a separate dichotomous item. 60% of the items for Windmill were either deleted as all incorrect or had flat item characteristics curves. This suggests that Windmill requires revision before use in future assessments.

**Treating indicators as polytomous items (sum of scores for indicator)**

Criteria not applicable to a task were not included in the final item set for Rasch analysis. As such, an indicator with 5 criteria, but only 4 applicable, had a maximum possible score of 4.

Similarly, an indicator with 5 criteria, but with one criteria having with all responses “not achieved”, a maximum score of 4 was possible. Criteria receiving all incorrect responses were included, but indicators with all responses “not achieved” were removed from the analysis (see Table 1.2):

- The Social management (SOM) a (communicate effectively) criteria for Lights Out Part A.
- The Self-management (SEM) a (Uses self-discipline and sets goals) criteria for Lights Out Part A
- SOM criteria a (communicate effectively) for Windmill.

Each participant was given a score for each indicator based on the sum of their scores of the relevant criteria within the indicator.

A large number (264 of 3133) of cases achieved extreme scores (either 0 or maximum possible score) for the indicators within tasks that were attempted/reached. This meant that item fit statistics were not available.

Summaries of fit statistics for each indicator can be found in the Excel workbook (See **Tab C** Excel workbook outline).

Category probability graphs for all remaining items are presented in **Tab D**.

### Treating indicators as polytomous items (Maximum score for indicator)

Criteria not applicable to a task were not included in the final item set for Rasch analysis. As such, an indicator with 5 criteria, but only 4 applicable, had a maximum possible score of 4.

However, an indicator with 5 criteria, but with one criteria having with all responses “not achieved”, a maximum score of 5 was possible if the fifth criteria was achieved, despite the all missing criteria earlier never being achieved.

Each participant was given a score for each indicator based on the highest criterion they achieved.

No indicators were removed by RUMM as extreme items.

Summaries of fit statistics for each indicator can be found in the Excel workbook (See **Tab C** Excel workbook outline).

Category probability graphs for all remaining items are presented in **Tab D**.

### Sum of scores vs Maximum score

- Sum of scores had a lower possible score
- Maximum score had zeroes for a number of categories
- Sum of scores had less proportion of disordered thresholds (65% compared to 90%).

## Preliminary wrap up: potential areas for further investigation

- Task comparisons, eg which tasks are easier, working better.
- Data treatment, eg Coding issues.
- Understanding data matrix and progression (dichotomous) this includes max and sum.

- There is currently no way of distinguishing between criteria within a task that a student has not reached and criteria a student has not attempted,
- Variability of the interactions between students.
- Polytomous
    - Looking for whether not additive as expected
    - Consistency with perceived conceptual leap, really know or don't know.
- Review the marking guide for all aspects of the task.
- Instances where students dropped out of the task before completing the majority of the task may affect the analysis of item quality and item function, particularly when students achieved the maximum or minimum scores for attempted task parts. An area for further investigation may be to select students based on the length of time spent in a task.

# Tab B: Summary tables and figures

**Table 1.1: Summary of Elements, indicators and number of criteria by task**

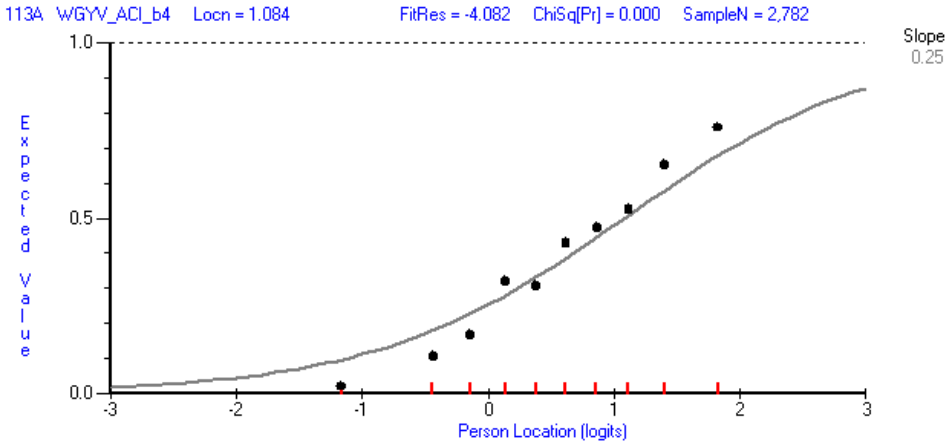
| Task | Part | Element | Indicator | Criteria |
|---|---|---|---|---|
| **Let's Tackle the Fish** | **A** | CI | b | 5 |
| | | SEM | a | 2 |
| | | | b | 5 |
| | | SOM | a | 4 |
| | **B** | CA | b | 5 |
| | | CI | b | 5 |
| | | SEM | a | 2 |
| | | | b | 5 |
| | | SOM | a | 4 |
| | **C** | CI | b | 4 |
| | | SEM | a | 3 |
| | | | b | 4 |
| | | SOM | a | 5 |
| | | | b | 5 |
| | **D** | CG | b | 4 |
| | | SEM | b | 5 |
| | | SOM | a | 5 |
| | | | b | 5 |
| **Lights Out v6** | **A** | SEM | a | 3 |
| | | SOM | a | 3 |
| | **B** | CG | a | 5 |
| | | SEM | b | 4 |
| | | SOM | a | 4 |
| | | | b | 4 |
| | **C** | CI | b | 5 |
| | | SOM | a | 5 |
| | | | b | 4 |
| | **D** | CA | b | 5 |
| | | CG | a | 5 |
| | | SOM | c | 4 |
| **Who's Got Your Vote** | **A** | CI | b | 3 |
| | | SEM | a | 3 |
| | | | b | 5 |
| | **B** | CA | a | 4 |
| | | CI | b | 5 |
| | | SEM | a | 3 |
| | | | b | 5 |
| | **C** | CG | c | 4 |
| | | SEM | a | 3 |
| | | | b | 5 |
| **Windmill** | **A** | CG | a | 3 |
| | | CI | a | 3 |
| | | SEM | a | 3 |
| | | SOM | a | 3 |
| | | | b | 3 |
| | **B** | CA | a | 3 |
| | | CI | a | 3 |
| | | SOM | a | 3 |
| | | | b | 3 |
| | **C** | CA | a | 3 |
| | | | b | 3 |
| | | SOM | a | 3 |
| | | | b | 3 |

**Table 1.2: Criteria deleted from dichotomous treatment**

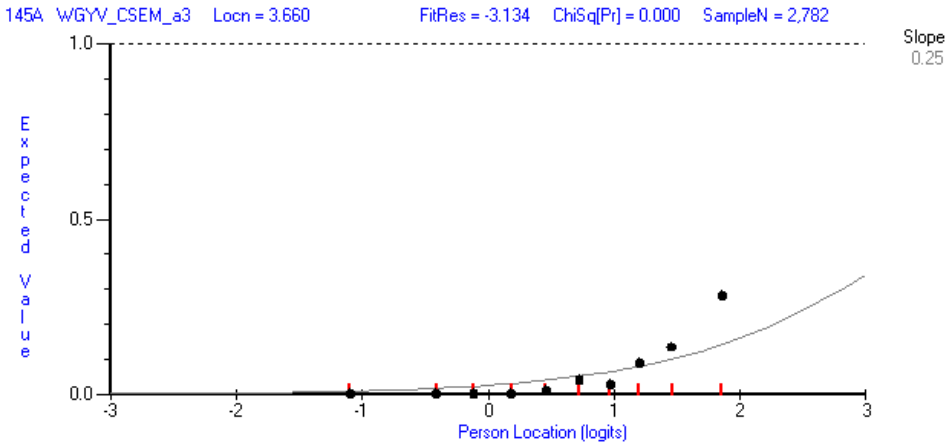
| Task | Part | Criterion | Criterion type (action/chat) | Removal reason |
|---|---|---|---|---|
| Lights Out v6 | A | SOM-a1 | Action | All Incorrect |
| Lights Out v6 | A | SOM-a2 | Action | All Incorrect |
| Lights Out v6 | A | SOM-a3 | Action | All Incorrect |
| Lights Out v6 | B | CG-a1 | Action | All Incorrect |
| Lights Out v6 | B | CG-a3 | Action | All Incorrect |
| Lights Out v6 | B | CG-a5 | Action | All Incorrect |
| Lights Out v6 | B | SEM-b1 | Action | All Incorrect |
| Lights Out v6 | B | SEM-b3 | Action | All Incorrect |
| Lights Out v6 | B | SEM-b4 | Action | All Incorrect |
| Lights Out v6 | B | SEM-b5 | Action | All Incorrect |
| Lights Out v6 | B | SOM-b1 | Action | All Incorrect |
| Lights Out v6 | B | SOM-b3 | Action | All Incorrect |
| Lights Out v6 | C | CI-b2 | Action | All Incorrect |
| Lights Out v6 | C | CI-b3 | Action | All Incorrect |
| Lights Out v6 | C | CI-b4 | Action | All Incorrect |
| Lights Out v6 | C | CI-b5 | Action | All Incorrect |
| Lights Out v6 | D | CA-b5 | Chat | All Incorrect |
| Let's Tackle the Fish | D | SOM-a5 | Chat | All Incorrect |
| Windmill | A | CI-a1 | Chat | All Incorrect |
| Windmill | A | SEM-a1 | Action | All Incorrect |
| Windmill | A | SEM-a3 | Action | All Incorrect |
| Windmill | B | CA-a1 | Chat | All Incorrect |
| Windmill | B | CI-a1 | Action | All Incorrect |
| Windmill | C | CA-a1 | Chat | All Incorrect |
| Windmill | C | CA-b3 | Action | All Incorrect |
| Windmill | C | SOM-a1 | Chat | All Incorrect |
| Windmill | C | SOM-a2 | Chat | All Incorrect |
| Windmill | C | SOM-a3 | Chat | All Incorrect |
| Windmill | C | SOM-b2 | Chat | All Incorrect |

## Item Characteristics Curve Exemplars

### Figure 1.1 “Increasing” Item Characteristics Curve




### Figure 1.2 “Increasing – hard” Item Characteristics Curve




### Figure 1.3 “Increasing – easy” Item Characteristics Curve

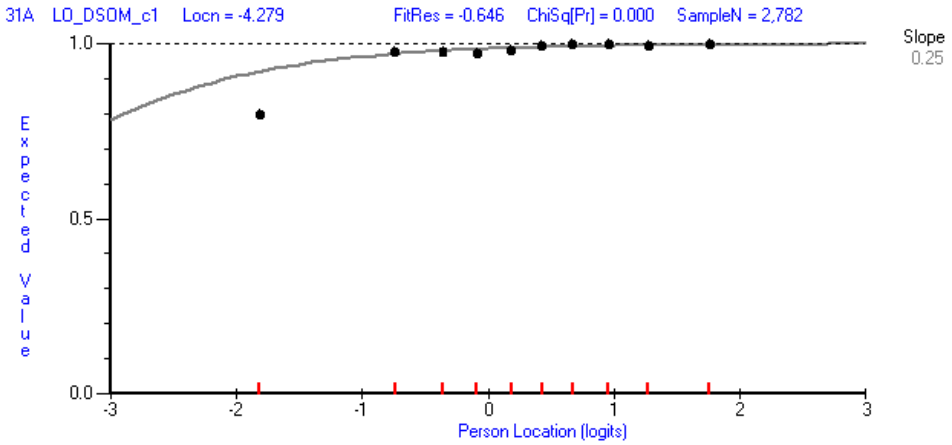


**Figure 1.4 “Increasing – steep” Item Characteristics Curve**

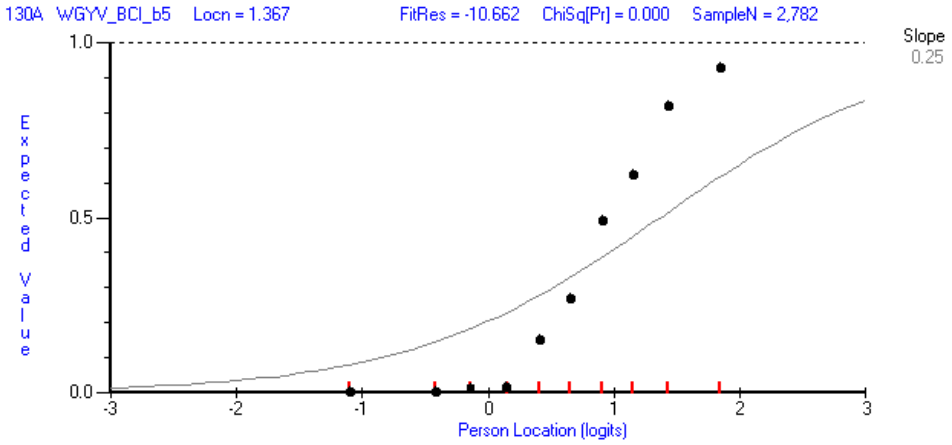



**Figure 1.5 “Flat – slightly increasing” Item Characteristics Curve**

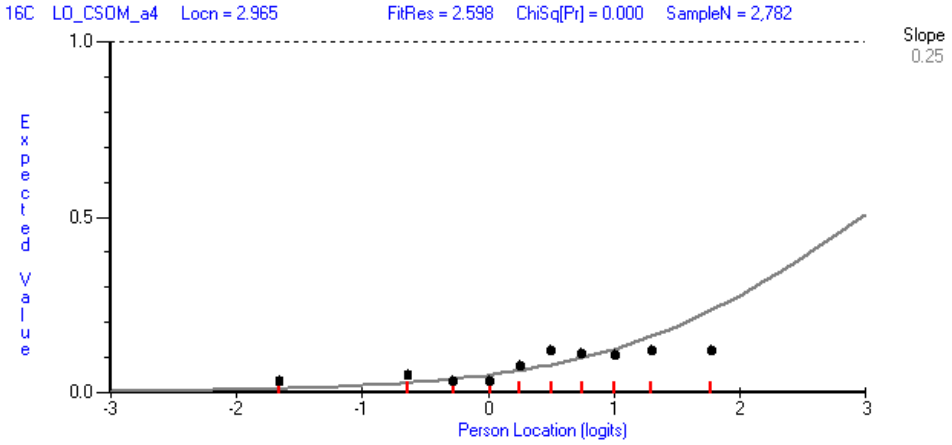



**Figure 1.6 “Flat” Item Characteristics Curve**

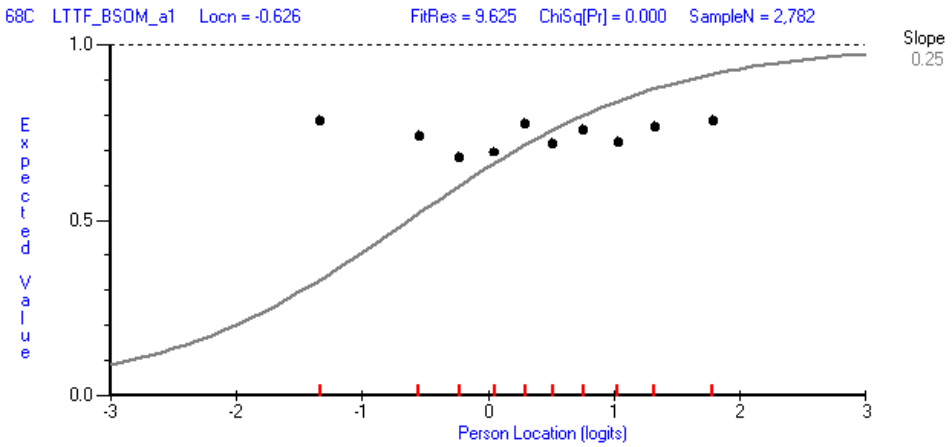


**Figure 1.7 “Flat – upper” Item Characteristics Curve**

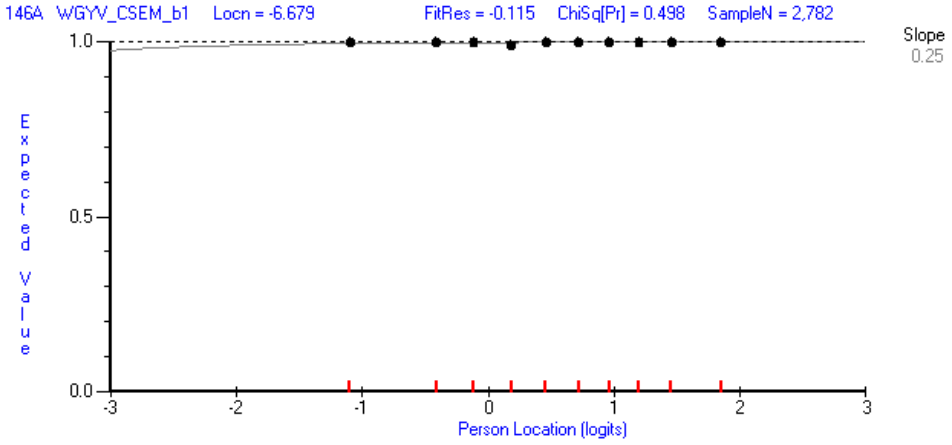
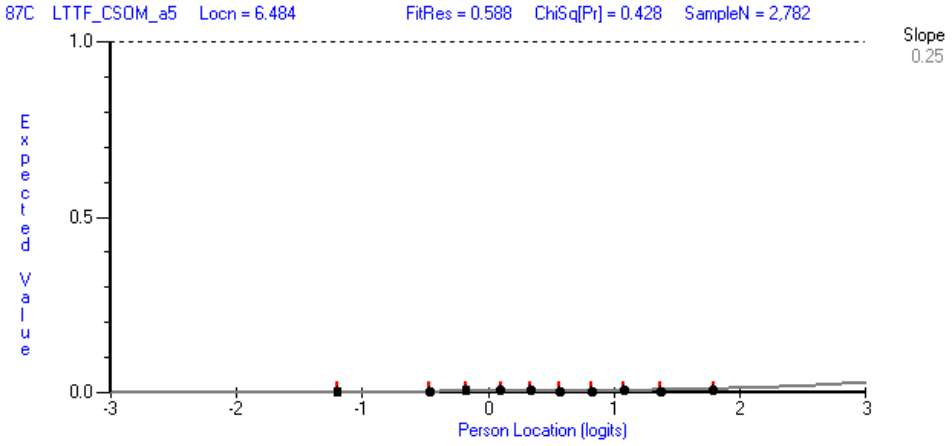



**Figure 1.8 “Flat – lower” Item Characteristics Curve**



## Item Characteristics Curves of individual criteria: Summaries by task and by overall assessment

**Table 1.3 Lights Out v6 task Item Characteristics Curve summary**

| Lights Out v6 | Part A | Part B | Part C | Part D | Task total | Percentage of task total |
|---|---|---|---|---|---|---|
| **Monotonically increasing items** | | | | | | |
| Increasing | 2 | 0 | 4 | 8 | 14 | 27.5% |
| Increasing - hard | 0 | 0 | 0 | 0 | 0 | 0.0% |
| Increasing - easy | 0 | 0 | 0 | 2 | 2 | 3.9% |
| Increasing - steep | 0 | 0 | 0 | 0 | 0 | 0.0% |
| Flat - slightly increasing | 0 | 5 | 2 | 0 | 7 | 13.7% |
| **Non-increasing items** | | | | | | |
| Flat | 0 | 0 | 0 | 0 | 0 | 0.0% |
| Decreasing | 0 | 1 | 0 | 1 | 2 | 3.9% |
| Flat - upper | 1 | 0 | 0 | 0 | 1 | 1.0% |
| Flat - lower | 0 | 2 | 4 | 2 | 8 | 15.7% |
| All incorrect | 3 | 9 | 4 | 1 | 17 | 33.3% |
| **Total Items** | 6 | 17 | 14 | 14 | 51 | |

**Table 1.4: Let's Tackle the Fish task Item Characteristics Curve summary**

| Let's Tackle the Fish | Part A | Part B | Part C | Part D | Task total | Percentage of task total |
|---|---|---|---|---|---|---|
| **Monotonically increasing items** | | | | | | |
| Increasing | 6 | 2 | 7 | 10 | 25 | 32.5% |
| Increasing - hard | 0 | 0 | 1 | 0 | 1 | 1.3% |
| Increasing - easy | 6 | 4 | 1 | 0 | 11 | 14.3% |
| Increasing - steep | 0 | 0 | 0 | 0 | 0 | 0.0% |
| Flat - slightly increasing | 1 | 0 | 2 | 3 | 6 | 7.8% |
| **Non-increasing items** | | | | | | |
| Flat | 0 | 1 | 2 | 2 | 5 | 6.5% |
| Decreasing | 0 | 0 | 1 | 0 | 1 | 1.3% |
| Flat - upper | 0 | 10 | 2 | 2 | 14 | 18.2% |
| Flat - lower | 3 | 4 | 5 | 1 | 13 | 16.9% |
| All incorrect | 0 | 0 | 0 | 1 | 1 | 1.3% |
| **Total Items** | 16 | 21 | 21 | 19 | 77 | |

**Table 1.5: Who's Got Your Vote task Item Characteristics Curve summary**

| Who's Got Your Vote? | Part A | Part B | Part C | Task total | Percentage of task total |
|---|---|---|---|---|---|
| **Monotonically increasing items** | | | | | |
| Increasing | 5 | 3 | 4 | 12 | 30.0% |
| Increasing - hard | 0 | 0 | 0 | 0 | 0.0% |
| Increasing - easy | 0 | 0 | 0 | 0 | 0.0% |
| Increasing - steep | 5 | 13 | 4 | 22 | 55.0% |
| Flat - slightly increasing | 0 | 0 | 0 | 0 | 0.0% |
| **Non-increasing items** | | | | | |
| Flat | 0 | 0 | 1 | 1 | 2.5% |
| Decreasing | 0 | 0 | 0 | 0 | 0.0% |
| Flat - upper | 1 | 1 | 1 | 3 | 7.5% |
| Flat - lower | 0 | 0 | 2 | 2 | 5.0% |
| All incorrect | 0 | 0 | 0 | 0 | 0.0% |
| **Total Items** | 11 | 17 | 12 | 40 | |

**Table 1.6: Windmill task Item Characteristics Curve summary**

| Windmill | Part A | Part B | Part C | Task total | Percentage of task total |
|---|---|---|---|---|---|
| **Monotonically increasing items** | | | | | |
| Increasing | 8 | 4 | 3 | 15 | 38.5% |
| Increasing - hard | 0 | 0 | 0 | 0 | 0.0% |
| Increasing - easy | 0 | 0 | 0 | 0 | 0.0% |
| Increasing - steep | 0 | 0 | 0 | 0 | 0.0% |
| Flat - slightly increasing | 0 | 0 | 0 | 0 | 0.0% |
| **Non-increasing items** | | | | | |
| Flat | 0 | 0 | 0 | 0 | 0.0% |
| Decreasing | 0 | 0 | 0 | 0 | 0.0% |
| Flat - upper | 3 | 1 | 1 | 5 | 12.8% |
| Flat - lower | 1 | 5 | 2 | 8 | 20.5% |
| All incorrect | 3 | 2 | 6 | 11 | 28.2% |
| **Total Items** | 15 | 12 | 12 | 39 | |

**Figure 1.9: Person-Item location map for individual criteria**

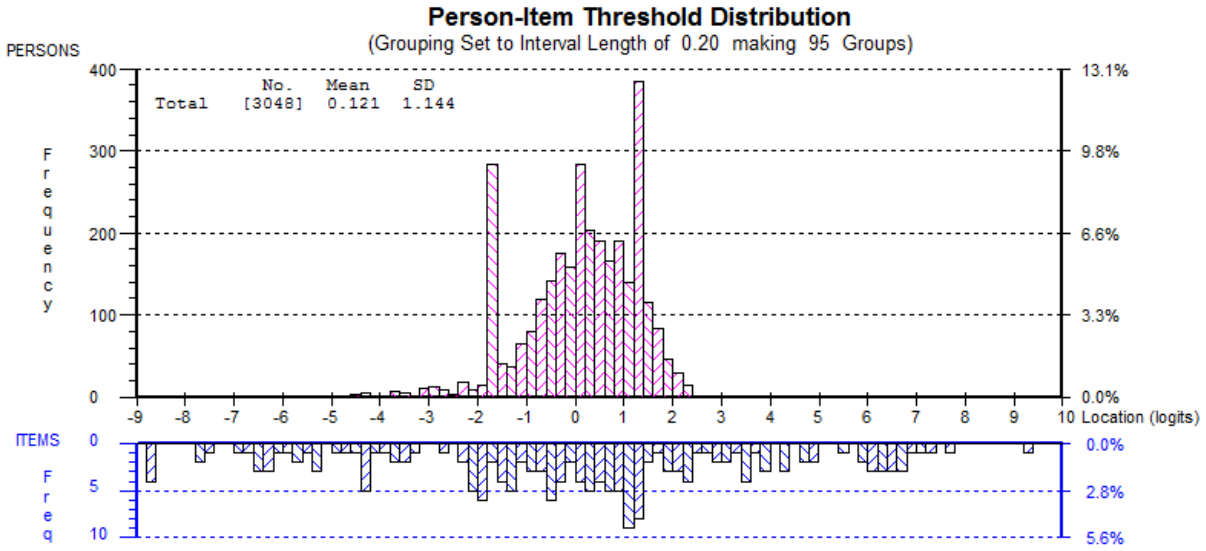


**Table 1.7: Entire assessment Item Characteristics Curve summary**

| | Lights Out v6 | Percentage of Lights Out v6 | Let's Tackle the Fish | Percentage of Let's Tackle the Fish | Who's Got Your Vote | Percentage of Who's Got Your Vote | Windmill | Percentage of Windmill | All tasks | Percentage of all tasks |
|---|---|---|---|---|---|---|---|---|---|---|
| **Monotonically increasing items** | | | | | | | | | | |
| Increasing | 14 | 27.5% | 25 | 32.5% | 12 | 30.0% | 15 | 38.5% | 66 | 31.9% |
| Increasing - hard | 0 | 0.0% | 1 | 1.3% | 0 | 0.0% | 0 | 0.0% | 1 | 0.5% |
| Increasing - easy | 2 | 3.9% | 11 | 14.3% | 0 | 0.0% | 0 | 0.0% | 13 | 6.3% |
| Increasing - steep | 0 | 0.0% | 0 | 0.0% | 22 | 55.0% | 0 | 0.0% | 22 | 10.6% |
| Flat - slightly increasing | 7 | 13.7% | 6 | 7.8% | 0 | 0.0% | 0 | 0.0% | 13 | 6.3% |
| Non-increasing items | | | | | | | | | | |
| **Flat** | 0 | 0.0% | 5 | 6.5% | 1 | 2.5% | 0 | 0.0% | 6 | 2.9% |
| Decreasing | 2 | 3.9% | 1 | 1.3% | 0 | 0.0% | 0 | 0.0% | 3 | 1.4% |
| Flat - upper | 1 | 2.0% | 14 | 18.2% | 3 | 7.5% | 5 | 12.8% | 23 | 11.1% |
| Flat - lower | 8 | 15.7% | 13 | 16.9% | 2 | 5.0% | 8 | 20.5% | 31 | 15.0% |
| All incorrect | 17 | 33.3% | 1 | 1.3% | 0 | 0.0% | 11 | 28.2% | 29 | 14.0% |
| Total Items | 51 | | 77 | | 40 | | 39 | | 207 | |

## Tab C: Table of contents of additional item level summary statistics provided in Excel file 'Item Statistics Summaries.xlsx'

**Data dictionary** – variable list and explanation of file contents

**Dichotomous** – Facility rate, fit statistics and item locations for dichotomous treatment

**Poly sum** – Facility rates and item locations for sum of scores polytomous treatment

**Poly max** – Facility rates, fit statistics and item locations for maximum score polytomous treatment

**ICC summaries** – Summary of item characteristics curves for dichotomous treatment by task, criterion and indicator

**Sum of score vs maximum score** – Comparison of highest response score and disordered thresholds for sum of scores and maximum score treatments

## Tab D: Table of contents of additional support files provided (e.g. code data, syntax)

**\RUMM ICCs and Category curves** – Item characteristics curves, item maps and category probability curves from RUMM

**\RUMM input files and RUMM files** – RUMM analysis files and intial input files

**\SPSS input and code –** Original data file used in analysis and SPSS syntax used to transform data.

# Appendix C: The marking/coding scheme and the data collected in detail

## Lights out

### Part A. Constructing a circuit

An individual student constructs the circuit by dragging and dropping the parts to the appropriate positions.

**The marking/coding for: SEMa**

| ***Self-management (SEM)*** | | | | | | |
|---|---|---|---|---|---|---|
| **Students set goals to work towards their achievement** | **Uses self-discipline and sets goals (a) A/C** | **Takes action** | **Follows simple instructions** | **Follows complex instructions** | **Makes a plan / sets goals** | **Analyses set goals** |
| | | Something moved to a location (SEMa1) | Drag components to correct locations – 4 actions (SEMa2) | Rotates components to make all parts connect – 4 actions (SEMa3) | | |

#### Actions

battery, lamp, setting box, switch [each placed in the correct location]

#### Positioning

switch, setting box [rotated to connect circuit], lamp and battery [not rotated]

**Comment**

Is this (SEM a) what is being assessed at this point; that is, the very first thing students are asked to do and do by themselves?

***The data collected***

The data for SEM a suggest that level 1 was very easy (-10.4), level 2 was relatively difficult (1.0), but level 3 (-0.6) was easier than level 2. The components must be placed in the circuit so that the current can flow. From observations of the first online trial, not all students immediately placed the components in the correct rotation, so it was difficult to draw clear conclusions from this data.

**The marking/coding for: SOMa**

| ***Social management (SOM)*** | | | | | | |
|---|---|---|---|---|---|---|
| **Coordinate and resolve potential differences in viewpoints, interests and strategies** | **Communicate effectively (a)**<br>**C** | **Attempts communication or**<br>**responds with little, irrelevant or incorrect information** | **Responds appropriately** | **Generates appropriate communication re context, self and others** | **Generates and responds to communication relevant to task requirements & constraints, clarifying problems** | **Initiates communication re deficits in common understandings** |
| | | Something typed in chat box but ***not using any predictive text*** (SOMa1) | Responds with appropriate greeting using ***Hello, hi, my name*** (SOMa2) | Generating communication ***how you, who you, where you, what you,*** *(SOMa3)* | | |

***The data collected***

All valid responses were coded as 0. This suggests that students did not respond in any of the ways described above or that their responses could not be coded.

## Part B. Simple brightest setting

**The marking/coding for: SEMb**

Students interact with each of the components available to them and use the chat room to determine the settings that produce the brightest light.

Student 1 can change the setting box and student 2 can change the switch only.

Students record their settings using two drop-down menus.

| ***Self-management (SEM)*** | | | | | | |
|---|---|---|---|---|---|---|
| **Students set goals and work toward their achievement** | **Work independently and show initiative (b)** | **Maintains a presence** | **Identifies a task** | **Attempts a task** | **Perseveres in a task** | **Completes a task successfully** |
| | | Interacts with their components (SEMb1) | | Both partners interact with components (SEMb3) | All settings were tried by both partners (SEMb4) | Identified correct combination of settings (SEMb5) |

***The data collected***

All valid responses were coded as 0. This suggests that students did not respond in any of the ways described above or that their responses could not be coded.

**The marking/coding for: SOMb**

| ***Social management (SOM)*** | | | | | |
|---|---|---|---|---|---|
| **Coordinate and resolve potential differences in viewpoints, interests and strategies** | **Work collaboratively (b) A/C** | **Is active in group activity** | **Responds to cues in information** | **Adapts activity based on contributions from others** | **Initiates and promotes interaction** |
| | | Changed settings on a component (SOMb1) | Tests a setting in response to a chat box suggestion (SOMb2) | Changed settings on both components (SOMb3) | |

***The data collected***

The data suggest that level 1 was very difficult (11.1), but level 2 was easy (-9.3); however, level 3 was difficult again (9.7). There were data recorded at level 4 even though no coding for that had been provided.

**The marking/coding of: SOMa**

| | ***Communicate effectively (a)*** ***C*** | ***Attempts communication or responds with little, irrelevant or incorrect information*** | ***Responds appropriately*** | ***Generates appropriate communication re context, self and others*** | ***Generates, and responds to, communication relevant to task requirements & constraints, clarifying problems*** |
|---|---|---|---|---|---|
| | | ***In the chat-box*** ***Hi, hello, who*** **(SOMa1)** | ***In the chat-box*** ***Its/switch up/down OR*** ***Its/switch top/bottom OR*** ***its/dimmer switch circle OR*** ***its/dimmer switch triangle OR*** ***its/dimmer switch square OR*** ***its/dimmer switch star*** **(SOMa2)** | ***In the chat-box*** ***Is it up/down OR*** ***Is it top/bottom*** ***OR*** ***Is it on/off*** ***Which/what shape*** ***OR*** ***Is it circle/triangle/square/ star*** ***(SOMa3)*** | ***In the chat-box*** ***Change switch OR*** ***Move up/down OR*** ***Move top/bottom OR*** ***Click circle/triangle/square/ star*** ***OR*** ***Try circle/triangle/square/ star/top/bottom*** ***(SOMa4)*** |

***The data collected***

The data suggest that level 1 was of medium difficulty (1.1), but level 2 was easier (0.1) and level 3 was more difficult (1.5), and level 4 was quite easy (-0.1). The writers anticipated that few students would demonstrate evidence for level 4, so it is a surprise that this appeared to be quite easy. It is difficult to explain why level 1 appeared to be difficult.

**The marking/coding of: CGa**

| Students generate ideas, possibilities and actions | Suggests actions and or ideas (a) A/C | Uses trial and error actions | Suggests an idea/ action to approach a task | Uses actions in logical/purposeful sequence | Compares ideas/actions as solutions to task | Uses a systematic approach to action sequence |
|---|---|---|---|---|---|---|
| | | Some settings were tried in each part (CGa1) | ***In the chat-box*** ***Check switch/ circle/triangle/square/ star*** (CGa2) | Switch turned on (CGa3) | ***In the chat-box*** ***Change switch/ circle/triangle/square/ star*** (CGa4) | Records correct combination of settings (CGa5) |

***The data collected***

The data suggests that level 1 was quite difficult (13.7), level 2 was very easy in comparison with level 1 (-8.3), level 3 was difficult again (9.1), and level 4 very easy (-10.3). It is important to note that levels 2 and 4 have criteria based on what students wrote in the chat box. There are no data recorded for level 5, which required students to record the settings. It is not known at this point whether no students entered any settings here or whether the data were not collected appropriately.

## Part C. Complex brightness settings

### The marking/coding: SOMb

Students use the chat box to share information about their settings with their partner as they can only see the settings on their own switch and settings box. From this, they need to determine:

i) which settings make the lamp go on

ii) which settings on the settings boxes make the lamp shine brightest.

| ***Social management (SOM)*** | | | | | | |
|---|---|---|---|---|---|---|
| **Coordinate and resolve potential differences in viewpoints, interests and strategies** | **Work collaboratively (b) A/C** | **Is active in group activity** | **Responds to cues in information** | **Adapts activity based on contributions from others** | **Initiates and promotes interaction** | **Makes proposals to benefit the group or assumes group responsibility** |
| | | Changed settings on a component (SOMb1) | Tests a setting in response to a chat box suggestion (SOMb2) | Changed settings on both components (SOMb3) | | *In the chat-box "we should/could …'*(SOMb5) |

***The data collected***

The data suggest that level 1 was not difficult (0.5); level 2 was a bit more difficult (1.8) – this may be a difficult response to collect data from, however, level 3 was quite easy (-2.7), which suggests that changing a setting once can then more easily be done on other components. Level 5 was quite difficult (4.0) – perhaps, students' conversation at this level did not occur very often. For example, in the case of one student instructing another, "we should" may not have been required. It is possible that this scenario of instruction occurred rather than a true collaboration of each partner contributing equally.

**The marking/coding: SOMa**

| | *Communicate effectively (a)* *C* | *Attempts communication or responds with little, irrelevant or incorrect information* | *Responds appropriately* | *Generates appropriate communication re context, self and others* | *Generates and responds to communication relevant to task requirements & constraints, clarifying problems* | *Initiates communication re deficits in common understandings* |
|---|---|---|---|---|---|---|
| | | *In the chat-box Hi, hello, who* **(SOMa1)** | *In the chat-box My switch up/down and My setting box circle, triangle, square, star* **(SOMa2)** | *In the chat-box Your switch up/down and Your setting box circle, triangle, square, star (SOMa3)* | *In the chat-box How dim/bright or When dimmest/brightest and When on/off (SOMa4)* | *In the chat-box What if (SOMa4)* |

***The data collected***

The data suggest that level 1 had some difficulty (2.5), level 2 was quite easy (-0.5), and level 3 easy (0.7), level 4 was difficult (2.5), level 5 was (1.2). This would suggest that what students iterate in the chat box (that is, how they communicate their thoughts about the problem) needs further research.

**The marking/coding: CIb**

| *Cognitive – inquiring (CI)* | | | | | | |
|---|---|---|---|---|---|---|
| **Students identify, explore and organise information** | **Collects and organises information (b)** **A** | **Uses an element of information** | **Uses isolated pieces of information** | **Uses relevant elements of information** | **Uses links in relevant information** | **Combines and uses relevant information from a number of sources** |
| | | Changes settings for one component (CIb1) | Changes settings on all component (CIb2) | Records correct combination of settings for switches (CIb3) | Produces the brightest lamp (CIb4) | records correct combination of settings for both components (CIb5) |

***The data collected***

The data suggest that level 1 was very difficult (10.3) and no further data were recorded for the other levels. Although these are similar actions to those required to turn the lamp on, it appears that to find the settings to make the lamp glow the brightest was far more difficult. This did require really good communication between the two students, as they needed to share information to achieve the result.

## Part D. Finding faults

**The marking/coding for: SOMc**

In this part, each student had two different light bulbs, a settings box and a switch. Students could drag and drop components into their section of the circuit (indicated by blue and red regions). Students needed to work systematically to determine which components are faulty and which are working, and drag them to the correct positions. For each person, one of the bulbs was broken and either the settings box or switch.

| *Social management (SOM)* | | | | | | |
|---|---|---|---|---|---|---|
| **Coordinate and resolve potential differences in viewpoints, interests and strategies** | **Make decisions (c)** | **Identifies needs** | **Identifies relevant criteria** | **Recognises or contributes ideas or results** | **Contributes to a group discussion** | **Evaluates the outcome of a decision** |
| | | Places one component in the circuit (SOMc1) | Tests all components for the circuit (SOMc2) | **In the chat-box *switch working/faulty/not working/broken* OR *dimmer switch working/faulty/not working/broken* OR *bulb/light working/faulty/not working/broken* (SOMc3)** | Correctly identifies their own faulty and working components **(SOMc4)** | |

***The data collected***

The data suggest that level 1 was relatively easy (-2.9), level 2 was harder (1.2), level 3 was slightly easier than level 2 (0.9), and level 4 was easier again (-0.9). This is very surprising as students found it very difficult to identify the faulty components during the trial.

***The marking/coding for: CAb***

| ***Cognitive – Analysing*** | | | | | |
|---|---|---|---|---|---|
| **Applies logic to developing a solution (b)** | **Takes action without knowledge or consideration of logical sequence** | **Takes an action considering the consequence** | **Takes a series of actions considering the consequences** | **Designs a logical course of action** | **Makes generalisations or evaluates plausible solutions** |
| | Tests some components (CAb1) | Correctly identifies one faulty component (CAb2) | Correctly identifies both faulty components (CAb3) | Correctly identifies faulty component for each person (CAb4) | **In the chat-box**<br>**If faulty/ not working/ broken then light won't glow/work/shine**<br>**(CAb5)** |

***The data collected***

The data suggest that level 1 was relatively easy (-3.1); level 2 was more difficult (1.3) – that is, identifying a faulty component; level 3 was slightly easier than level 2 (0.8) – that is, identifying more than one faulty component, and even easier to identify the faulty components of your partner level 4 (-0.8), which requires communication. No data were provided for level 5.

**The marking/coding for: CGa**

| Students generate ideas, possibilities and actions | Suggests actions and or ideas (a) A/C | Uses trial and error actions | Suggests an idea/ action to approach a task | Uses actions in logical/purposeful sequence | Compares ideas/actions as solutions to task | Uses a systematic approach to action sequence |
|---|---|---|---|---|---|---|
| | | Puts own component into position but no light bulb (CGa1) | Puts 2 components into position but no light bulb (CGa2) | Tests own component using a light bulb in circuit (CGa3) | ***In the chat-box Must/need light*** (CGa4) | Tests all components using working light (CGa5) |

***The data collected***

The data suggest that level 1 was quite difficult (4.7), whereas level 2 was easier (-4.5), and level 3 was even easier (-1.7); however, level 4, which required communication in the chat box, was difficult (3.3); and level 5 that required the testing of all components was easier (0.6). Again, our experience in the first trial strongly suggested that very few students were able to correctly test and identify the faulty components, limiting student access to other levels. How the data were collected at higher levels needs further enquiry.

# Appendix D

## Collaborative Assessment Alliance: Ireland

Significant changes in society require learners to have a wide, adaptive knowledge base and understanding to enable them to be active participants in the communities in which they live and work. The premium in today's world is not merely on students' acquiring information, but on their ability to analyse, synthesise, and apply what they have learned to address new problems, design solutions, collaborate effectively, and communicate persuasively (Pellegrino, 2014). Recent education reforms aspire to embed key competences in teaching and learning through rich learning outcomes. What is less clear is how existing assessment methods can properly evaluate skills such as critical thinking, problem solving, creativity, communication and collaboration.

No single assessment can evaluate all kinds of learning, rather, a coordinated system of assessment is needed that incorporates the assessment of higher order skills, includes real world skills of collaboration and communication, and engages students in instructionally valuable activities (Darling-Hammond et al., 2013). They argue that students need to engage in tasks that measure these complex skills and not evaluate a proxy for these skills.

Jenkins et al. (2006), acknowledge the digital and participatory worlds that young people need to negotiate. However, student engagement within these worlds is often blurred by the notion of the student as a 'digital native' and by technical and technological approaches in schools that replicate traditional methods of assessment and instruction as opposed to embracing new ones (Claxton, 2007). In this traditional format, assessment tends to be associated with institutions and sanctioned assessors, whereas Gee (2010) argues that it has a natural home in human action and learning. This human action now includes interaction with technology and in his thoughts on Actor Network Theory, Latour (2005), places objects and non-human entities on an equal footing, and states that technology and social practices are inextricably linked. Lakhana (2014) agrees with this notion that we cannot separate technology from its social relations, as people are co-constructors of knowledge. The increasing influence of digital worlds means that young people are seen to be taking on new participatory and collaborative roles in learning online and outside the classroom, and there is a growing interest in incorporating these roles and practices inside education.

This notion of the social and collaborative context of assessment is explored currently through the Collaborative Assessment Alliance project (http://www.caa21.org/), where students are assessed on their ability to collaborate on social and cognitive domains through the medium of online synchronous collaborative tasks. The local alliance in

Ireland is managed by the National Council for Curriculum and Assessment (NCCA), a statutory body with the vision for leading innovation in education. The NCCA supports innovation in schools by engaging with learners, teachers and parents through undertaking, using and sharing research as a basis for advice and debate on education. The CAA was seen as a key opportunity to promote and provoke debate on collaborative problem solving while working closely with teachers on learning design and implementation.

Thirteen schools from around Ireland developed digital synchronous collaborative tasks that challenge students to solve problems through collaborating with their partner. The teachers in these schools have explored the research and theories on collaborative assessment, and their implication for classroom and school practice. In designing their tasks, the teachers have considered how the content and the context of the lesson relates to the development of skills and how conceptual and metacognitive knowledge is built. They have further considered how the collected assessment data can be evaluated using a progression framework that identifies the student's participation and collaboration in the task on both cognitive and social domains. The skills that are targeted in the tasks align closely with the key skills of junior cycle and senior cycle education in Ireland, and recognises the role that digital learning can play in the development of these key skills. Collaborative problem-solving tasks that are mediated through a digital platform will support and enhance these key skills and enable students to flourish in an uncertain and challenging future.

It was agreed that the project would focus in on the curricular area of science, as the specifications in this subject area at junior and senior level were nearing completion. Science teachers were invited to apply to participate in the initiative through an online expression of interest form, resulting in 20 teachers who have committed to be participants in the Irish branch of this international alliance. National and international experts in curriculum and assessment design delivered a series of workshops as part of the teacher professional learning. The content of the workshops covered a wide range of topics that focused in on exploring the:

- socio-cognitive approaches to assessment
- localisation and personalisation of assessment
- nature of collaborative assessment
- design of collaborative learning tasks
- design of rubrics to measure collaborative engagement with a task

The science teachers received expert tuition and support in the areas outlined above to assist them in preparing for the design and creation of collaborative tasks that were trialed in schools across the country.

The tasks are unique in that they provide a window into how students collaborate, and how they approach collaboration; how they set their learning agenda. The professional development that teachers have recieved as a result of participation in the project has encouraged them to consider the role that digital learning has in assessment of not only the cognitive ability of students but also their social and collaborative abilities.

The indicative outcomes of the Irish research include capacity building in the understanding and implementing of digital assessment, something underexplored in the Irish context. The initiative has resulted in the development of expertise in collaborative problem solving and performance assessment, an approach that is better suited to measuring higher order skills (Pecheone et al., 2010). A model for teacher professional development in assessment of collaboration and higher order skills has been developed. There is now research evidence to support the design of tasks that validly assess collaboration and problem solving.

## References


Claxton, G. (2007) *'Expanding young people's capacity to learn'*. British Journal of Educational Studies, 55 (2), pp. 115-134.

Darling-Hammond, L., Herman, J., Pellegrino, J., (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Gee (2010). Human Action and Social Groups as the natural home of assessment: Thoughts on 21st Century Learning and Assessment. In Shute, V. and Becker, B. (Eds.) *Innovative Assessment for the 21st Century.* (pp. 13 – 39). Springer, New York.

Jenkins, H., Clinton, K., Purushotma, R., Robison, A.J. and Weigel, M. (2006) *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Available from: http://digitallearning.macfound.org/atf/cf/%7B7E45C7E0-A3E0-4B89-AC9C-E807E1B0AE4E%7D/JENKINS_WHITE_PAPER.PDF

Lakhana, A. (2014). *What is Educational Technology? An Inquiry into the Meaning, use and Reciprocity of Technology*. Canadian Journal of Learning and Technology. 40 (3)

Latour, B. (2005). *Reassembling the Social*. Oxford: Oxford University Press.

Pecheone, R., Kahl, S., Hamma, J., Jaquith, A. (2010). *Through a Looking Glass: Lessons Learned and Future Directions for Performance Assessment*. Stanford, CA: Stanford University.

Pellegrino, J. W. (2014). *Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress*. Keynote address in Proceedings of the 2014 Conference of the International Association for Educational Assessment, Singapore