

# Assessing Graduate Screen Production Outputs in Nineteen Australian Film Schools

Josko Petkovic

## Abstract

Can the output of creative practices, such as screen production, be measured or is it in the eye of the beholder? This paper presents the results of the Australian Learning and Teaching Council (ALTC) funded Priority research project *Assessing Graduate Screen Production Outputs in Nineteen Australian Film Schools*. The aim of the project was to show that assessment of image-based creative works is as consistent as assessment conducted in traditional discipline areas. Thirty (30) screen production academics from 22 institutions were tested using the same sample of 45 short honours productions and a carefully selected set of 34 criteria. Creative works alone were assessed without any written components or exegeses. Rasch psychometric modeling was then used to test the internal consistency of the assessment data. The results confirmed the hypothesis that screen production assessors are consistent for all criteria considered. There was also little difference in the assessments of “objective” and “subjective” criteria which invites a host of questions regarding the status of subjective and objective verification procedures within creative arts. Is this distinction still valid in the psychometric Google universe?

## 1. Historical Context

**The Medium of Film:** For the purpose of this study the medium of film can best be defined as a time-based medium made up of images and sounds. It is a comparatively recent invention that is generally considered to have come into existence on 28 December 1895 with the screening of the first Lumiere documentary footage. In the early days of its discovery film was considered to be a duplicating medium something akin to the closed circuit television (CCTV) recording today. For this reason it was also considered to be most suitable for use as a scientific medium and most useful for recording time-based events and experiments. However, it was not long before it was discovered that film’s time-based nature gave it a great narrative capacity that could convey powerful emotions. Within twenty years of film’s inception the first features were being produced and this narrative form has dominated the medium ever since.<sup>1</sup>

**Television:** In Australia, a major change took place in September 1956 with the launch of a television broadcast station in Sydney. From that time onwards film and television have co-existed together, at times competing with one another, at times complementing one another, but always informing one another. For the purpose of this project filmmakers and TV producers will be considered as being interchangeable within the term “screen producers”.

**The First Training Institution in Australia:** Until the launch of television in 1956, filmmakers in Australia were created on the job by the industry itself. There were no screen production programs in the university sector although films were discussed in academia within humanities, literature and philosophy programs – but usually in relationship to their literary inspirations. However within the first ten years of the introduction of television in Australia, the industry grew to number 73 television stations, supported by over 200 recognized advertising agencies and some 60 film producing agencies. The need of the industry gave rise to 1966 Weeden Report which investigated the possibility of establishing film and television training in educational institutions.<sup>2</sup>

The first film and television training program in Australia was established in 1966 as a Diploma of Art (film and television) within the Department of Graphic Design in the Swinburne Arts School, Swinburne Technical College, Melbourne.

**Emergence of Screen Production as Art:** Swinburne Film and Television School went on to be one of the most influential screen production institutions in Australia. Given its influence it is worth noting that the origin of the School in the School of Art was not accidental. Barbara Paterson in her book *Renegades: Australia's First Film School from Swinburne to VCA* describes the reasons for the art-based foundations of the School as prescribed by its founder Brian Robinson:

The film and television course was established in the Art School because of the generally accepted principle that film and television were fundamentally visual arts. The idea was to train artists in the mechanics of film and television. Not only was the course physically located in the art building, but the underlying philosophy and assumptions were to remain those of the Art School for many years.

Paterson then goes on to quote some of Robinson's artistic philosophies:

Contemporary film and television afford the artist a flexible new vehicle for expression ... in which all the arts, it would seem, seek to speak with one voice. It is the dominant voice of the twentieth century.<sup>3</sup>

Swinburne's example was followed in other states – mostly in the art-based technical institutions and colleges of advanced education (CAE). The high point of this development was in 1972 with the establishment of the first national film school in Sydney. Given decades of pent up creative energy prior to its establishment it was not surprising that the Australian Film and Television School (AFTS) quickly blossomed.<sup>4</sup> The teaching in schools extended to master level as was the case with most other film schools in the world.

The principle “film is art” was still in place in 1992 when the Swinburne Film and Television School became incorporated into Victorian College of the Arts, The University of Melbourne. The same art-based “accepted principle” was also inscribed in the range of arts funding bodies such as the Australian Council for the Arts which would not fund anything that had to do with education or research scholarship.

The inverse form of this art-based “accepted principle” was also inscribed in the funding guidelines of principle research funding institutions, namely the Australian Research Council (ARC). Just as the Australian Council for the Arts (OzCo) would not fund anything “educational”, so too the Australian Research Council regulations stipulated that they do not fund anything artistic:

*ARC Discovery Projects* does **NOT** support the following work:

6.5.1b. activities leading solely to the creation or performance of a work of art, including visual art, musical compositions, drama, dance, designs and literary works, for which Commonwealth Government support is provided through the Australia Council for the Arts.<sup>5</sup>

At the time when the Australian Film and Television School was established creative arts were not recognized as an ARC research category and screen production undergraduates did not go on to do honours or PhD programs. Master of Arts was considered to be the appropriate terminal program for the practicing artists. We can thus observe in passing that in Australia in the late 1970s screen education seemed neatly divided into two streams: The best image-based practitioners went to arts training programs, did Master of Arts and then sought funding from the Australian Council for the Arts; whereas screen scholars completed PhD programs, most probably in Humanities, and then sought funding from the Australian Research Council.

**From Art to Scholarship:** A major cultural shift in this art–scholarship divide began to take place during 1970s as the result of the growing influence of television. Visual culture was no longer something that was carried out in rarefied atmosphere of high art. Visual signs were everywhere, as ubiquitous as television itself. This cultural shift was accompanied by the growth of screen and televisual programs in tertiary institutions. These were based on the simple proposition that communication took place across many codes and visual codes were arguably becoming just as important as the code of language.<sup>6</sup> The new perspective, in turn, supported the proposition that screen production could be a scholarly activity as well as an art practice. To a large extent the question of whether film was either art or scholarship was no longer considered relevant – at least within this emerging sector of the academia.

**Theory/ Practice:** The transformation of image-making from art to scholarship initially took place within departments of humanities and literature where issues related to the nexus of theory and practice were also being championed. The growing popularity of these programs created momentum of its own and it was inevitable that this intellectual development would express itself through practical programs. The trend accelerated during the period of the so called Dawkins Revolution (1987–1992) when many Australian technical institutions that once were traditional training grounds for filmmakers, namely institutes of technologies and colleges of advanced education, were transformed into universities by the intervention of the federal government.<sup>7</sup>

**Assessing Screen Production – Art or Scholarship?:** Screen production programs are now commonplace in Australian tertiary institutions. They are a major component of what is broadly known as the creative arts sector within Australian universities. As indicated earlier, this sector has grown rapidly in the last 30 years. Yet this rapid growth has not been accompanied by an adjustment of academic regulations, which for the most part continue to be based on the established paradigms of scholarship and on empirical, scientific and written conventions.

Screen Production remains a problematic activity within academia. Indeed, the tradition of film as art did not go away just because semiotic signs became popular in some sections of academia. Artistic activity is still considered by many academics as being essentially sensual, subjective and not all that scholarly.

**Contested Paradigms of Scholarship:** Academia today could be described as having two strands of scholarship: the established strand of conventional scholarship and the emerging strand of creative arts scholarship. On the conventional scholarship side there is an increasing acceptance that practice-based creative arts scholarship is a valid way of approaching certain research questions. Similarly, the creative arts sector has gone some way towards accommodating conventional scholarship methodologies.

The coexistence of these two strands of scholarship within academia still leaves some unresolved questions. For example, conventional scholarship has reasonably articulate, predictable, objective and verifiable methodologies. Can we say something similar for the methodologies of the creative arts sector? Can these methodologies be described, measured, prescribed? How can such work be evaluated? What are the observables, the evidence and verification processes of creative methods? What is the role of subjectivity, emotion, sensuality, audience and impact in these verification processes? Similar questions to these have been asked before and Aristotle's *Nichomachean Ethics* provides a good point of entry to the work of past scholars.<sup>8</sup>

**Ethical and Practical Knowledge:** For Aristotle, art and scholarship were considered as different forms of the same intellectual activity. In *Nichomachean Ethics*, he divides knowledge into two types: the first type is the pure theoretical type of knowledge of the kind we encounter in the hard sciences; the second type of knowledge is the practical and ethical knowledge of the kind one encounters in politics, life sciences and creative arts. For Aristotle the ethical knowledge is the higher order of knowledge since it is self evident that politics (and political knowledge) decides what sciences (and theoretical knowledge) is embraced by society. Aristotle makes another important distinction between the two types of knowledge: theoretical knowledge tends to be precise in the way that science and mathematics are precise; in contrast ethical knowledge is not precise in the same way that democracy is not precise. It could be said that we still follow an Aristotelian hierarchy of knowledge since we generally endorse democratic values above all others. I also contend that the screen production paradigm of scholarship has much in common with the Aristotelian ethical and practical scholarship.

**Psychometrics:** We can end this short historical reflection with an additional observation. For over four hundred years creative arts had been relegated to the margins of scholarship mostly because art-based methodologies tended to be complex, spontaneous (practice-based), subjective, sensual, unpredictable, difficult to

describe and difficult to measure. Often the most interesting, the most sophisticated and the most nuanced creative works were dismissed as too vague and too personal because they did not operate on the linear logic of most scientific explanations. As a result, the most important outcomes of creative arts were often relegated to magical formulas and divine religiosity rather than to the highest attribute of human endeavour.

We are fortunate to be living at a time when complex phenomena are not set-aside in such a dismissive manner but are becoming increasingly accessible to measurement. Creative artists should welcome this type of development as it has the potential to again confirm the high status of creativity. This project makes use of one such measuring procedure to shed light on the nature of creative arts practice.

Specifically this paper presents the results of the Australian Learning and Teaching Council (ALTC) funded Priority research project *Assessing Graduate Screen Production Outputs in Nineteen Australian Film Schools*. The aim of the project was to show that assessment of image-based creative works is as consistent as assessment conducted in traditional discipline areas. To confirm this hypothesis thirty (30) screen production academics from 22 institutions were tested using the same sample of 45 short honours productions and a carefully selected set of 34 criteria. Creative works alone were assessed without any written components or exegeses.

Of the 30 screen production academics that were tested 25 were from Australian – Western Australia (WA), Victoria (VIC), New South Wales (NSW) and the Australian Capital Territory (ACT), Queensland (QLND), South Australia (SA) – and 5 were from the United Kingdom (UK) (for full details of collaborators see Appendix 1). The makeup of the six panels, each consisting of 5 academics, is depicted in Figure 1 below.

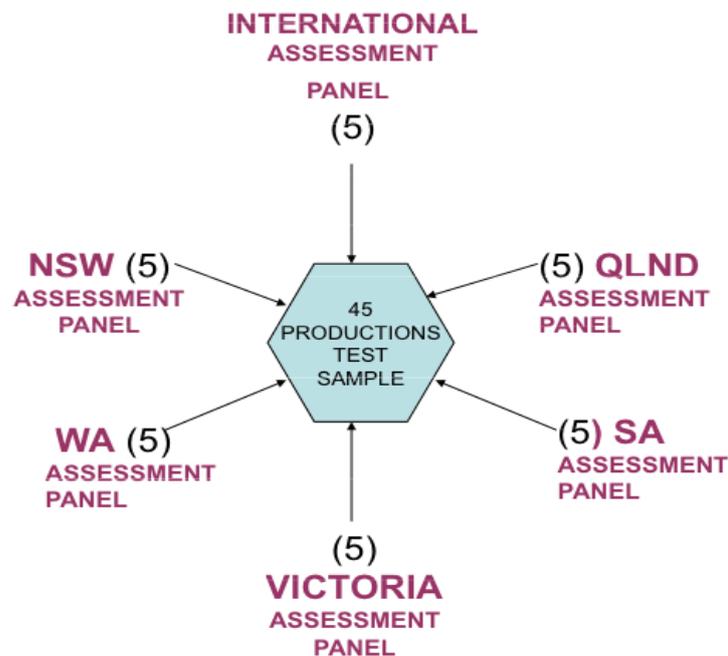


Figure 1 Project Assessment Scheme

If consistency of these 30 academics could be confirmed it would signify that the activity of screen production was based on some predictable and potentially scholarly procedure and not on some kind of unsubstantiated subjective judgment that resides in the eye of the beholder.

## 2. Experimental Design

**Internal Consistency – Rasch Analysis:** The analysis in this project takes up Rasch psychometric modelling which tests the internal consistency of the assessment data.<sup>9</sup> Rasch analysis does this by breaking up the overall assessment task into many smaller assessment components/ criteria and then comparing each component with the overall assessment mark.

The details of this analysis are somewhat complex and will be greatly simplified in the description that follows. It is, however, sufficient at the moment to remind ourselves that every screen production text presents a viewer with a conglomerate of assessable codes and components that are seemingly independent such as camera, sound, editing, script, performance, direction, set design and music, to note a few. At the same time each one of these components contribute in some way to the final percentage mark.

**Component Scale:** To carry out Rasch analysis with such components parts it is also necessary to arrange an appropriate assessment scale for each component. Ideally these smaller assessment components need to comply with a shared understanding of assessment criteria and scale that are used when “writing” with images. The “publication” criteria used in this project is a good case in point as indicated below:

<b>PUBLICATION</b>	
<b>Q4</b>	<b>THE PUBLICATION VALUE OF THIS PRODUCTION FOR ITS PROJECTED AUDIENCE IS:</b>
Very low	0
Modest	1
Moderate	2
High	3
Very High	4

If we have an appropriate set of such criteria, each with an individual assessment scale, we have a scheme in which the total assessment score at any one time can be compared with each component score. The scheme of this kind gives us a handle on what is being assessed when an overall mark for a production is given. Ideally, the Total (%) Mark and each component mark should behave in a similar fashion: increases in the Total (%) Mark should be reflected in an increase of the component. Developing a set of criteria that relate positively to the Total (%) Mark is the key to a successful Rasch analysis. Selecting these criteria was a major outcome in this project and is summarised below.

**Identifying Component Criteria:** What is a suitable set of scholarly criteria for assessing screen productions beyond the overall percentage (%) score? A myriad of possibilities offer themselves for consideration. If we consider each creative work as a communication between the originator of the production (sender) and the audience (receiver) many fields of study are relevant to the assessment and to the assessment criteria including Communication Studies (message), Media Studies (medium), Cultural Studies (culture), Textual Studies (text), Discourse Analysis (discourse), Design Studies (set design), Media and Marketing Studies (exhibition and distribution) along with Gender Studies, Social Science, Anthropology, Ethnography, History, Environmental Studies, Biology, Zoology, Medicine, Legal Studies and Comparative Literature, to name a few. Each one of these fields of study comes with its own set of assessment criteria and there appears to be no limit of criteria that can be invoked. Clearly, however, having too many criteria is counterproductive as the assessment process could go on forever. Ideally, the criteria set should be necessary to the assessment process and in that sense be a complete criteria set. Also, the assessment scale should be comparable with the assessment scale of other components and it should be an even scale. Furthermore, the component criteria should be different and not repeat one another (non-redundant). Thus, in a summary form, the criteria set should be:

- (i) necessary
- (ii) complete
- (iii) evenly graded (linear)
- (iv) independent/ non-dependent

**Dependency:** Of all the above requirements it is the last requirement (iv) of non-dependence that is philosophically the most curious. The schematic diagram of assessment in Figure 9 suggests that assessment components are linked together like branches on a tree and any one component is not independent of other components. Given this, it is difficult to understand how one can isolate assessment of any one component without implicating the quality of all other components in the assessment process. However we also do know that this is commonly done and we are quite used to individual prizes being awarded for direction, cinematography, and so on. It seems that from an assessment point of view assessment criteria can be both dependent and independent. This presents us with a logical dilemma that must be resolved before the final assessment criteria set is selected. How, then, can we explain this ability to assess seemingly interconnected qualities in an independent fashion?

**Gestalt Levels:** Gestalt theory tells us that when some things combine together a new quality emerges from the component parts that was not evident initially. In some sense, this new gestalt quality is independent of the constituent parts and cannot be recognized by analysing the parts alone. The novelty of this emerging quality is well captured by the gestalt proclamation: “the whole is greater than the sum of the parts”.<sup>10</sup>

If this emerging gestalt quality is not to be found in the individual component parts, we then need to consider it as existing at another logical level. Accordingly, these different logical levels can be considered as independent of one another if they are kept separate from one another. The process of gestalt emergence can occur from level to level giving the following range of emerging gestalt qualities:

<b><u>Gestalt Level</u></b>	<b><u>Content</u></b>
Level 1	Component Parts
Level 2	Gestalt of (Component Parts)
Level 3	Gestalt of (Gestalt of (Component Parts))
Level 4	Gestalt of (Gestalt of (Gestalt of (Component Parts)))
.....	.....
Level n	Gestalt of (Gestalt of (Gestalt of (Gestalt ... (Component Parts)...)))

The best way of selecting an appropriate set of assessment criteria is to (i) work along a particular component level until all assessment criteria at that level are exhausted and then (ii) repeat the whole process again on the next level, and then on the next level until the whole latent quality to be assessed is captured.

**Subjective/ Objective:** Because screen production is such a hybrid medium, these component criteria descriptions must necessarily include both “subjective” and “objective”, logical and phenomenological elements. The following levels of quality were considered when selecting the assessment criteria for this project:

Level 1	<b>Quantity of work:</b> number of shots, scenes, number of actors, extent of work (30 second ad, short, documentary, feature, series, sequels).
Level 2	<b>Process – component parts and crew functions:</b> Concept, script, research, direction, camera, editing, sound, music, CG Effects, animation, art design, performances, locations.
Level 3	<b>Gestalt of parts:</b> the first gestalt triad – topic, media and art
Level 4	<b>Gestalt enhancement:</b> textual connectivity between: topic, media, art, and timing.
Level 6	<b>Contextual Social Gestalt:</b> Social context, cross-referencing and relevance.
Level 7	<b>Gestalt Verification:</b> Clarity, intellectual value, believability, emotive level.
Level 8	<b>Emerging qualities:</b> Originality, innovation
Level 9	<b>Audience appeal:</b> exhibition potential, publication value.
Level 10	<b>Ethical values:</b> As apprehended by the audience.
Level 11	<b>Overall mark (%)</b> .

**Screen Production Assessment Scale (SPAS):** After much consideration a set of 34 criteria was selected and codified as 22 questions with corresponding assessment scales. These 22 questions were purposefully not ordered in any systematic fashion although the levels outlined above are reasonably easy to discern. There were also two additional questions (Q10 and Q19) not related to the assessment criteria and not included in the criteria list below:

**SCREEN PRODUCTION ASSESSMENT CRITERIA AND SCALE**

- Q1 OVERALL MARK (%)**
- Q2. STRONG AND SUSTAINED CONTRIBUTION TO THE QUALITY:**
- |    |             |        |
|----|-------------|--------|
| Q2 | Concept     | (0, 1) |
| Q2 | Script      | (0, 1) |
| Q2 | Research    | (0, 1) |
| Q2 | Direction   | (0, 1) |
| Q2 | Camera      | (0, 1) |
| Q2 | Editing     | (0, 1) |
| Q2 | Soundscape  | (0, 1) |
| Q2 | Music       | (0, 1) |
| Q2 | CG Effects  | (0, 1) |
| Q2 | Animation   | (0, 1) |
| Q2 | Art Design  | (0, 1) |
| Q2 | Performance | (0, 1) |
| Q2 | Locations   | (0, 1) |
- Q3 AUDIENCE:** This Production will be best appreciated by:  
General (1), Specialist (1), Both (2), Neither (0).
- Q4 PUBLICATION:** The publication value of this production for its projected audience is: 0, 1, 2, 3, 4.
- Q5 ORIGINALITY:** The level of originality in this production is: 0, 1, 2, 3.
- Q6 CLARITY:** The clarity of the theme/ central concept is: 0, 1, 2, 3.
- Q7 ARTISTIC QUALITY:** The artistic quality of this production is: 0, 1, 2, 3.
- Q8 ENHANCEMENT:** Different elements in this production (camera, casting, editing, sound etc) broadly enhance one another: 0, 1, 2, 3.
- Q9 INSPIRATION:** This production inspires thoughts of other narratives, other references and other contexts: 0, 1, 2, 3.
- Q11 GESTALT:** The best element of this production is:  
What: its content (1), How the content is presented (1)  
Both: what it has to say and how it says it (2)  
Neither what it has to say nor how it says it (0).
- Q12 INTELLECTUAL:** The intellectual level of this production is: 0, 1, 2, 3
- Q13 TIMING:** The duration of this production seems: Long (0), Short (0), Appropriate (1), Perfect (2).
- Q14 PRODUCTION:** Overall this production is:  
Imitative (0), Superficial (1), Ordinary(2), Thoughtful (3), Insightful(4).
- Q15 EMOTIVENESS:** Emotive level of this production is: 0, 1, 2, 3.
- Q16 BELIEVABILITY:** Believability of this production is: 0, 1, 2, 3.
- Q17 SOCIAL RELEVANCE:** Social relevance of this production is: 0, 1, 2, 3.
- Q18 PRODUCTION VALUE:** Overall production values are: 0, 1, 2, 3.
- Q20 ETHICS:** The Use of emotions in this production is: Inappropriate (0), Appropriate (1), Exhilarating (2), Cathartic (3), Not Applicable (0).
- Q21 INNOVATION:** The level of innovation in this production is : 0, 1, 2, 3.
- Q22 AIMS:** This production has attained its projected aims: 0, 1, 2, 3.
- Q23 WORK EVIDENT:** The amount of work evident in this production is: 1, 2, 3.
- Q24 EXHIBITION:** Your recommend exhibition site for this productions is:
- |                         |     |
|-------------------------|-----|
| Specialist Conference   | (6) |
| Theatrical Distribution | (5) |
| TV Broadcast            | (4) |
| International Festivals | (3) |
| Local Festivals         | (2) |
| YouTube                 | (1) |
| Institutional Archives  | (0) |

The above criteria components constitute the Screen Production Assessment Scale and the major outcome of this project. If the maximum component values are added together the sum comes to 70. It is possible to use the SPA scale in the same way as a percentage scale, as long as it is remembered that the range of the SPA scale is 0 to 70 rather than 0 to 100 per cent. We broadly expect that the sum total of SPAS marks for each production will correlate well with the overall percentage mark.

**Data Set:** The above 34 criteria set was used by 30 thirty assessors to assess 45 short honours productions giving rise to approximately 45,900 items of data.

**Rasch Analysis:** Once the above data set was in place it was then tested for internal consistency using Rasch statistical analysis. The analysis itself was carried out by researchers from the Pearson Psychometrics Laboratory (UWA): specifically by Professor David Andrich, Director of Pearson Psychometric Laboratory, and by Associate Professor Irene Styles, Senior Research fellow with the Pearson Laboratory.<sup>11</sup>

**The Software Used In The Analysis:** The RUMM 2030 software by Andrich, Sheridan and Luo was used to analyse the data.<sup>12</sup> This software provided an extensive range of information for assessing the quality of items in the SPA scale including several different statistical and graphical tests of fit between the data and the model. This information was used to establish an overall conclusion about the quality of the SPA scale. The results that follow are extracts from the Pearson Psychometric Laboratory analysis report commissioned for this project.

### 3. Results

**Correlation Between SPAS And Total % Scores:** On Pearson's correlation coefficients scale (0 to 1) the correlations between the mean SPA score and the overall mean % score of the 45 productions and 30 assessors was 0.973.<sup>13</sup> This means that the two were highly correlated with each other – as one would expect if these two assessments aim to represent similar characteristics of the productions.

**The Production Separation Index (PSI):** How much overlap/ separation is there in the criteria set? The Production Separation Index (PSI) for this data was 0.958 (on a scale of 0 to 1.0) which means that the reliability of the scale was very good. The Pearson Psychometric Laboratory report noted some instances of criteria dependencies indicating one or other of the following pairs of items may be redundant:

- Q7(*artistic quality*) with Q8(*elements enhance each other*);
- Q7(*artistic quality*) with Q18(*overall production values*);
- Q5(*originality*) with Q21(*innovation*);
- Q18(*overall production values*) with Q23(*amount of work evident*);
- Q16(*believability*) with Q17(*social relevance*).

These dependencies have the effect of artificially raising the Production Separation Index (PSI) a little, but their effect is not likely to have been marked. Accordingly the report recommended that all these items be retained.

**Visual Diagnostics:** Arguably the most interesting aspect of the Rasch analysis is the graphic depiction of the results which provide immediate and illuminating diagnostics of internal consistency or a lack thereof. The remainder of this paper will focus on this element of the analysis and specifically on the characteristic assessment curves derived from the analysis of assessment responses. These curves provide an elegant support for the consistency of the assessment hypothesis and need only a short statistical introduction.

**Rasch Ideal Paradigm:** The Rasch analysis begins with the construction of an ideal positive correlation model for the “Total SPAS Mark” and the “Component Mark” as depicted below in Figure 2.

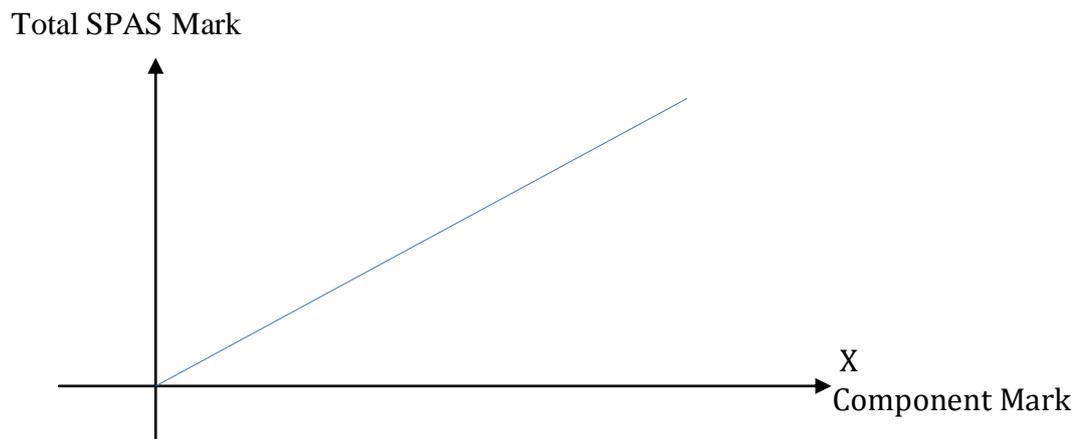


Figure 2 Positive correlation between Total SPAS Mark (Y) and Component Mark (X)

This simple depiction of correlation is adjusted to account for large numbers of data items by changing the X-axis into a logarithmic “logits” scale. The logarithmic scale compresses the amount of data into a usable space on the x-axis. This, in turn, changes the ideal straight line correlation in Figure 2 into an ideal wave-type of curve as depicted in figure 3 below.

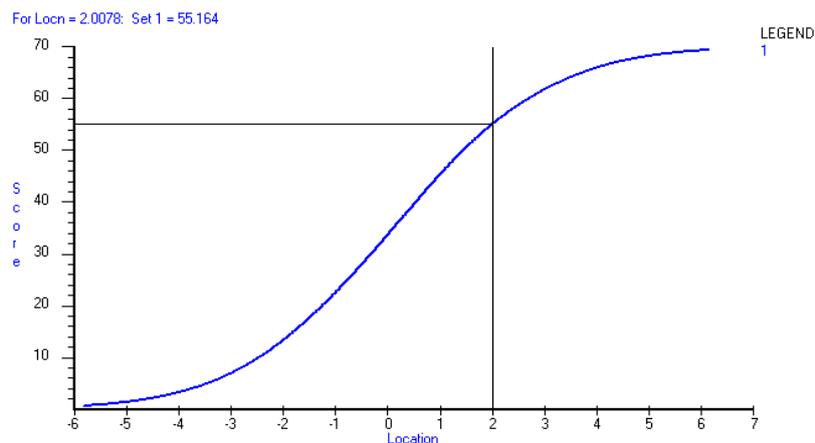


Figure 3 Equating graph for raw total scores and transformed scores in logits on the SPAS

It should be noted that the most accurate reading of the assessment grades is around the middle of the characteristic curve namely at zero logits on the x-axis.<sup>14</sup>

**Ideal Visual Comparison:** The approach in the Rasch analysis is that the data fit the Rasch ideal model. The comparison between the ideal and the actual criteria component is then visually evident as the difference between the ideal criteria assessment curve (line) and the actual assessment data curve (dots) as depicted in the diagrams that follow. Three types of comparisons with the ideal (line) curves were typically found in this project.

The first of these is the excellent fit of data to the ideal characteristic curve as in Figure 4:

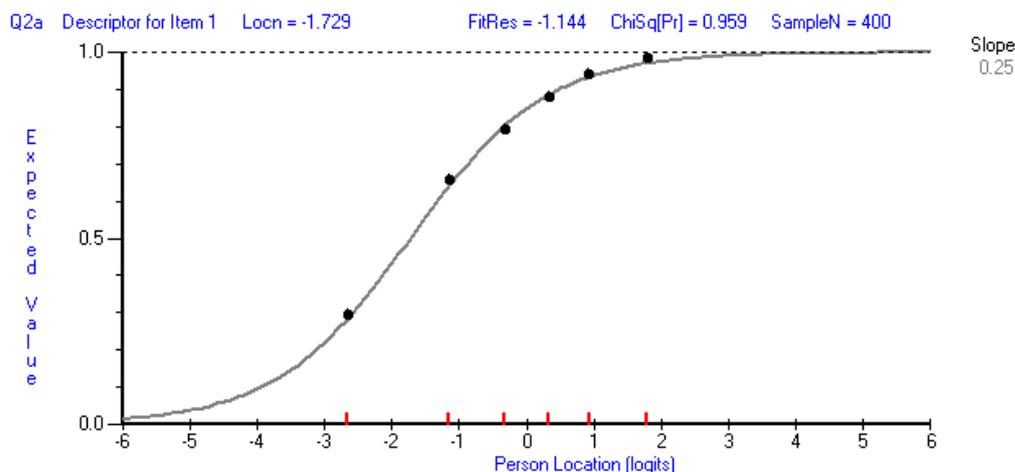


Figure 4 Excellent fit of the data (dots) to the ideal curve (line)

The second type of comparison is of data in Figure 5. This criteria item tends to under-discriminate a little (the obtained curve is a little flatter than the expected curve). This means that highest score groups are tending to score less than expected relative to lowest groups who are scoring a little higher than expected.

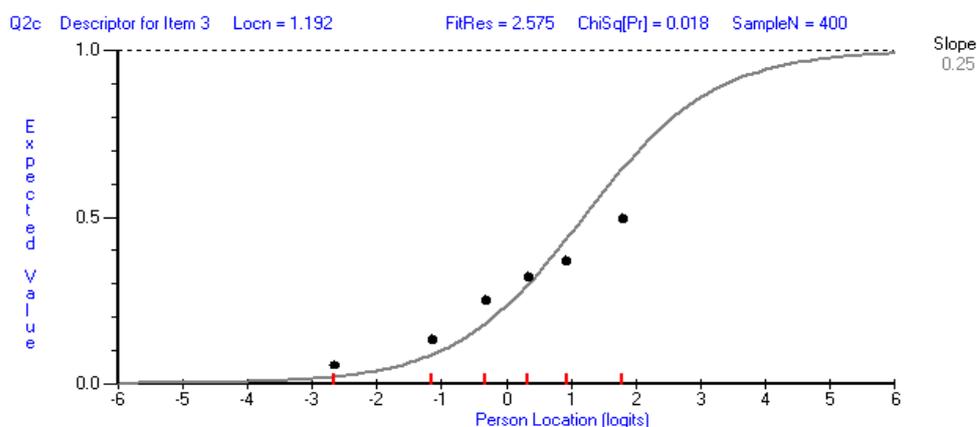


Figure 5 Under-discrimination of the data (dots) to the ideal curve (line)

The third type of comparison depicted in Figure 6 below is one that tends to over-discriminate across total score groups. This means that lower total score groups tending to score less than expected and higher total score groups tending to score higher than expected.

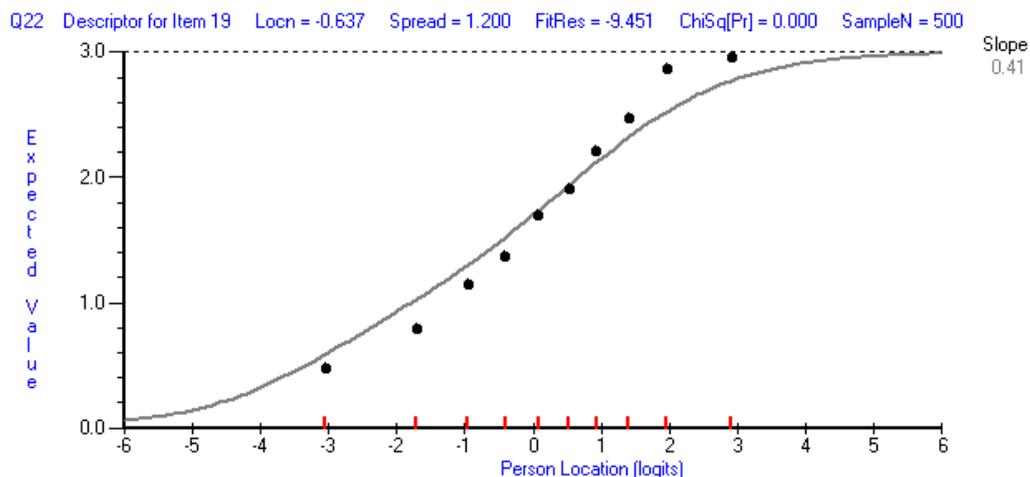


Figure 6 Over-discrimination of the data (dots) to the ideal curve (line)

### Graphs (Item Characteristic Curves or ICCs) of all items

The characteristic curves responses depicted below are accepted as fitting the Rasch model well, thus indicating internal consistency amongst the set of items which comprise the SPA scale.<sup>15</sup> The diagnostic comments associated with the curves below have been created mostly by the Rasch analysts. They are inserted here to help understand the graphs and are not significant enough to undermine the overall consistency of the items overall in measuring the quality of productions.

## Q2 CONCEPT

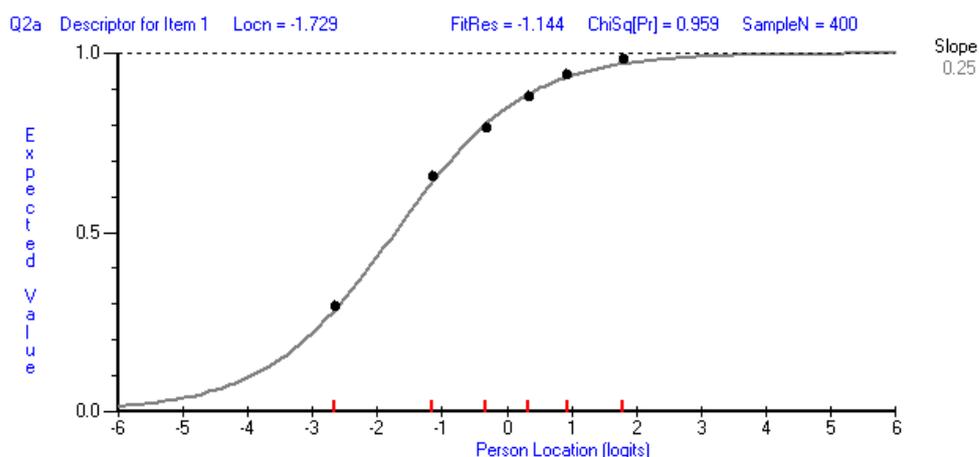


Figure 7 ICC of Q2 (*Concept: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item is an excellent fit to the ideal Rasch “Q2 Concept” curve (line).

## Q2 RESEARCH

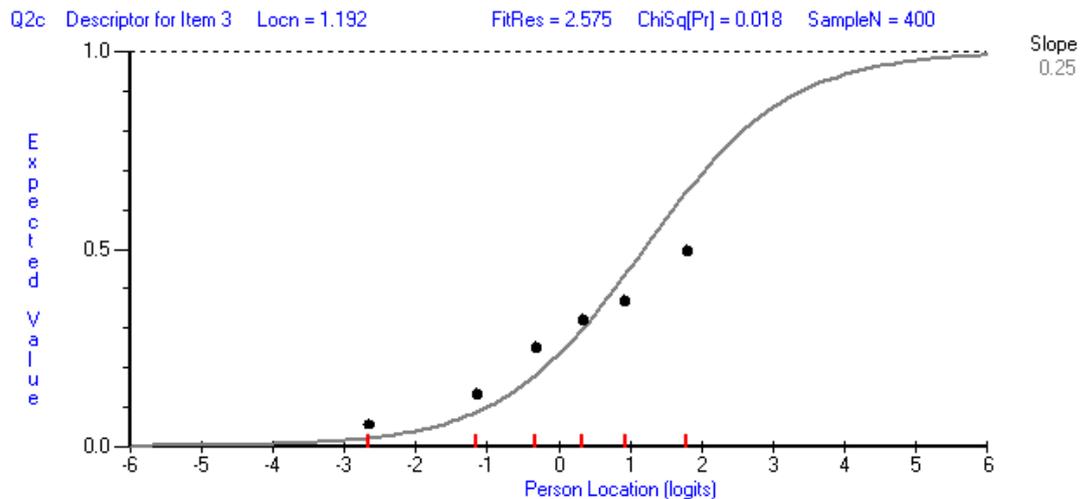


Figure 8 ICC of Q2 (*Research: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item tends to under-discriminate a little – the obtained curve is a little flatter than the expected from the ideal curve (line).

## Q2 SCRIPT

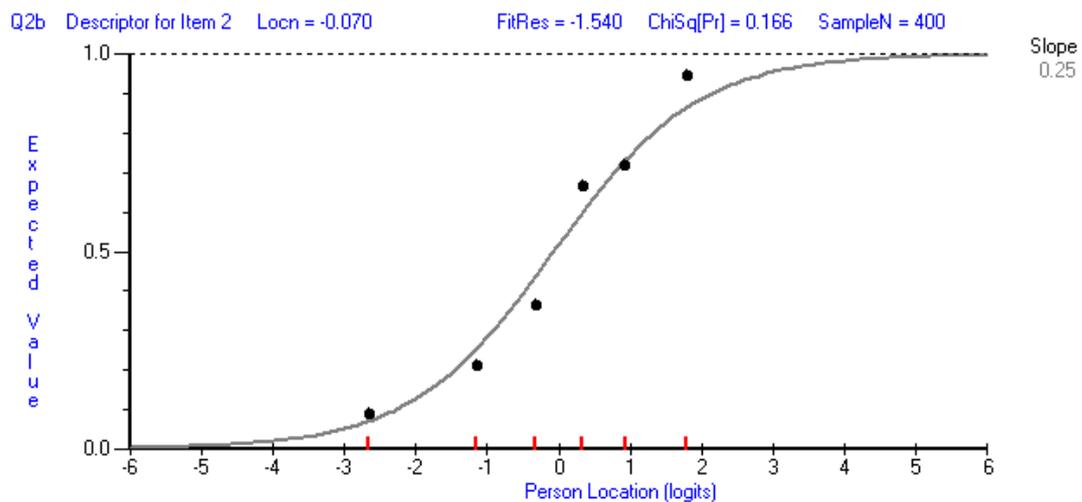


Figure 9 ICC of Q2 (*Script: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item is a good fit with some over-discrimination to the expected ideal curve (line).

## Q2 DIRECTION

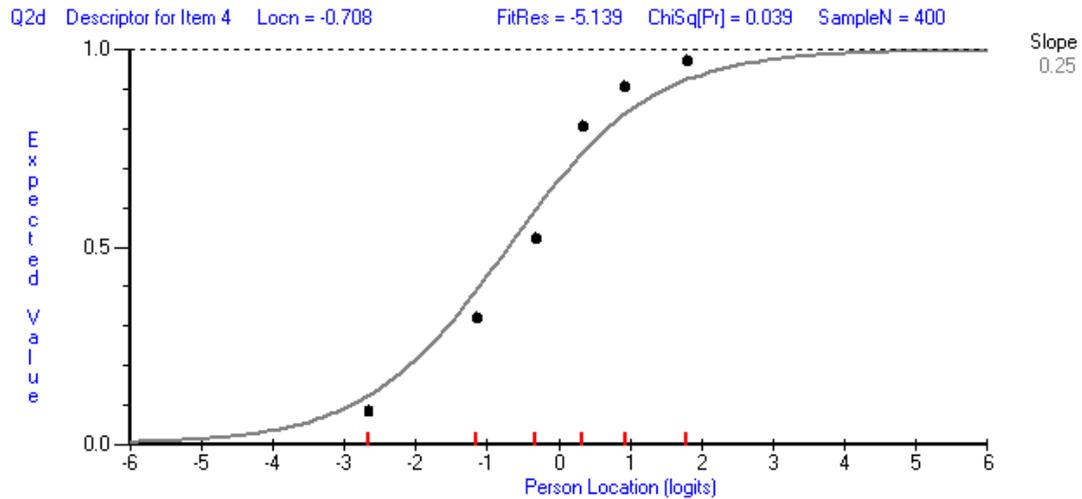


Figure 10 ICC of Q2 (*Direction: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item is a good fit with some over-discrimination to the ideal curve (line).

## Q2 CAMERA

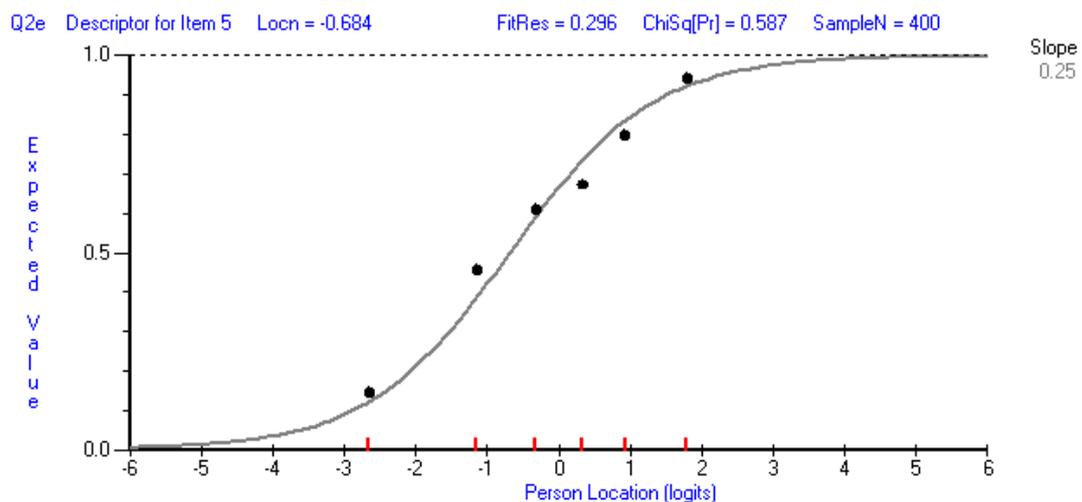


Figure 11 ICC of Q2 (*Camera: Strong and sustained contribution to....*)

Comment: The assessment data for this criteria item (dots) is a good fit to the ideal curve (line).

## Q2 EDITING

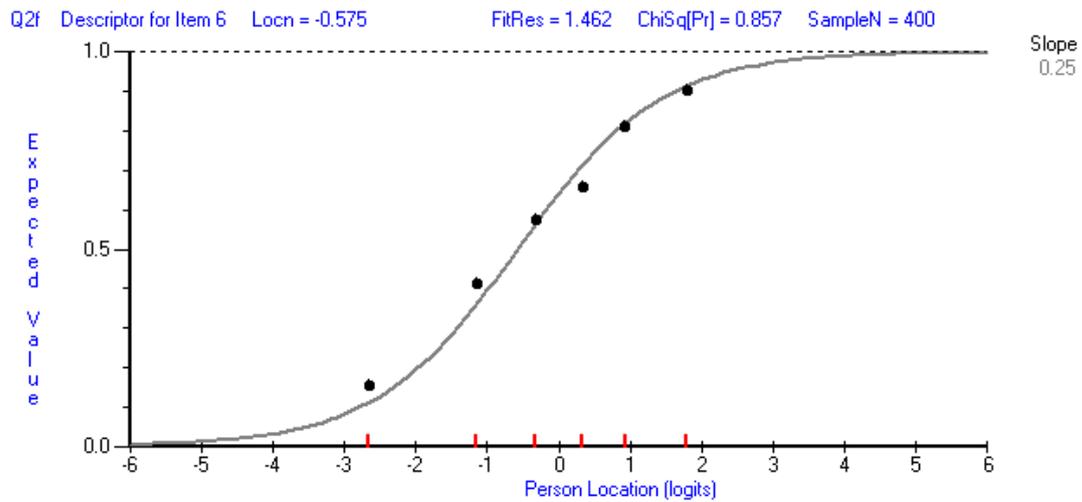


Figure 12 ICC of Q2 (*Editing: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item is a good fit to the ideal curve (line).

## Q2 SOUNDSCAPE

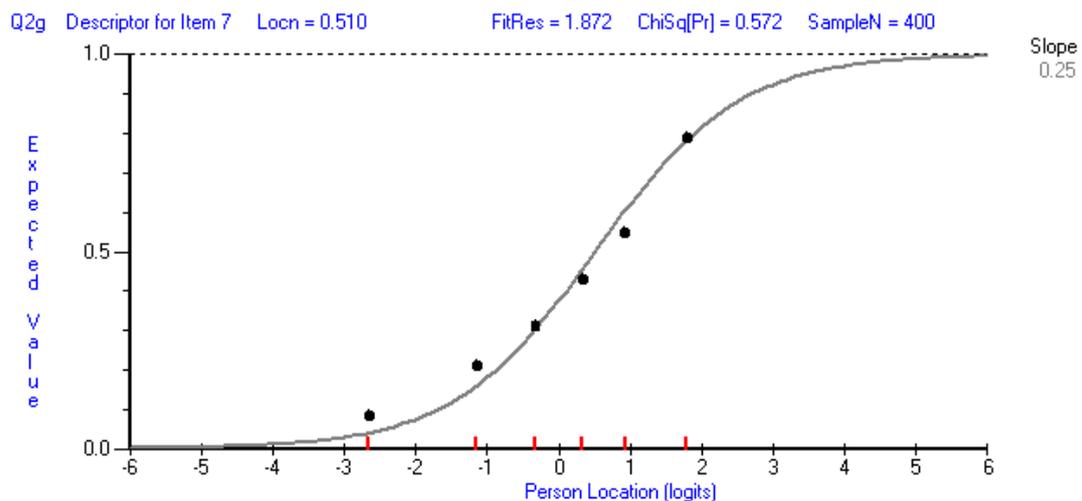


Figure 13 ICC of Q2 (*Soundscape: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item is a good fit to the ideal curve (line).

## Q2 MUSIC

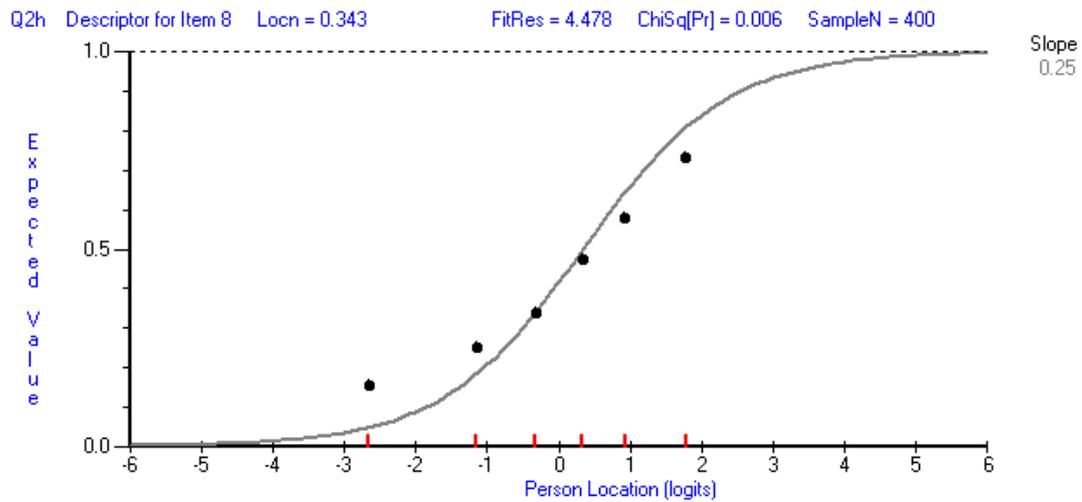


Figure 14 ICC of Q2 (*Music: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item tends to under-discriminate a little – the obtained curve is a little flatter than the expected curve (line).

## Q2 PERFORMANCE

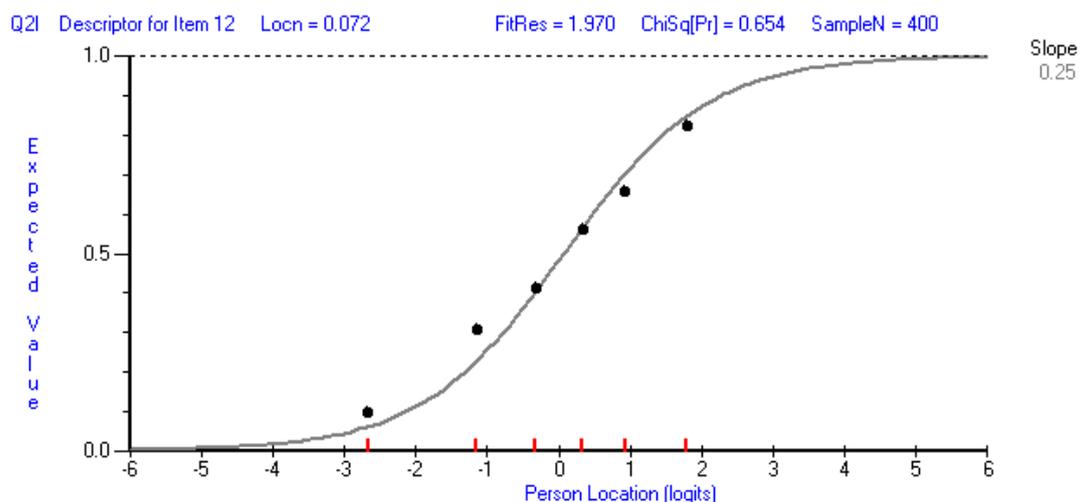


Figure 15 ICC of Q2 (*Performance: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item is mostly a good fit to the ideal curve (line) – slight under-discrimination.

## Q2 LOCATIONS

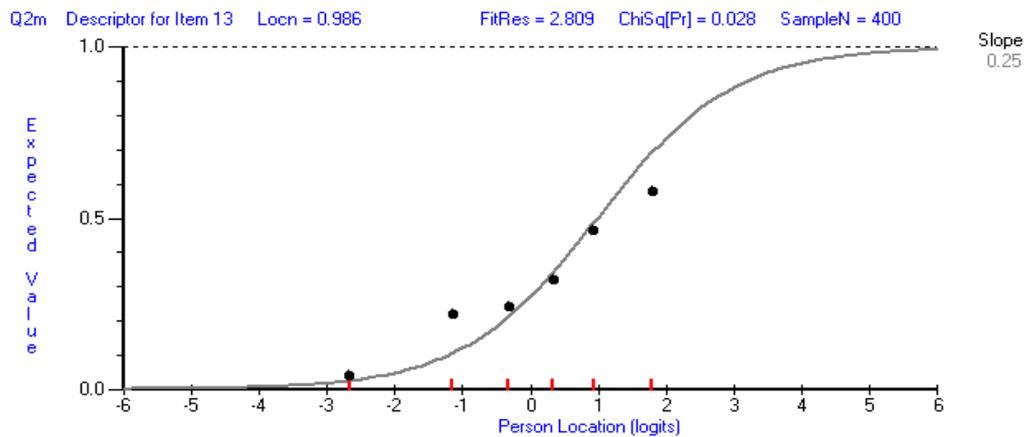


Figure 16 ICC of Q2 (*Locations: Strong and sustained contribution to....*)

Comment: The assessment data (dots) for this criteria item tends to under-discriminate a little – the obtained curve is a little flatter than the expected curve (line).

## Q3 AUDIENCE

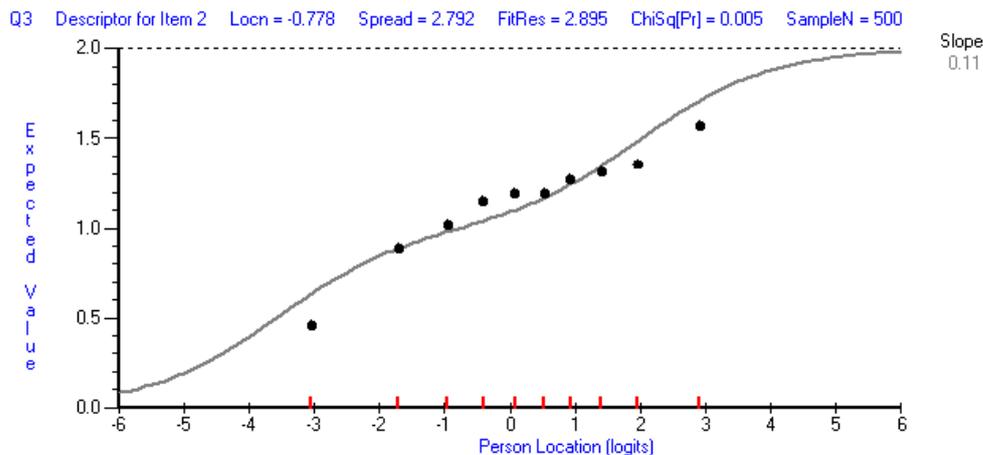


Figure 17 ICC of Q3 (*Audience: This production will be best appreciated....*)

Comment: The assessment data (dots) for this criteria item tends to under-discriminate a little from the middle to the upper total score groups and tends to over-discriminate between the lowest and middle total score groups. This means that highest score groups are tending to score less than expected relative to middle groups who are scoring a little higher than expected. And the lowest total score groups are scoring a little lower than expected relative to the middle groups who are scoring a little higher than expected from the model. The “wobbly” look of the expected curve is due to the location of the criteria scale thresholds which tend to be unequal in their relative placements.

## Q4 PUBLICATION VALUE

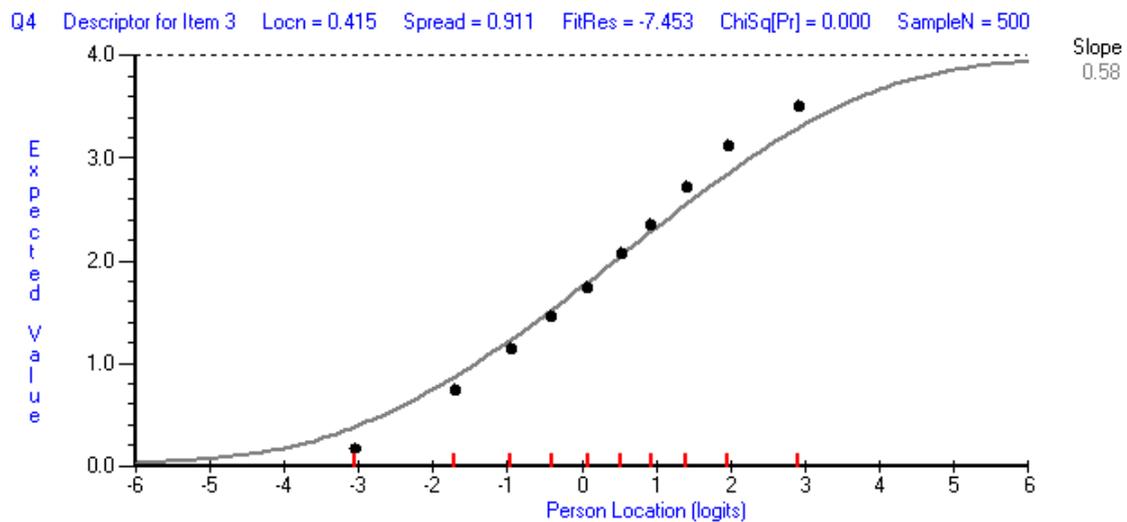


Figure 18 ICC of Q4 (*The publication value of this production for its projected audience....*)

Comment: The assessment data (dots) for this criteria item is tending to over-discriminate a little (lower total score groups tending to score less than expected and higher total groups tending to score a little higher than expected (line)).

## Q5 ORIGINALITY

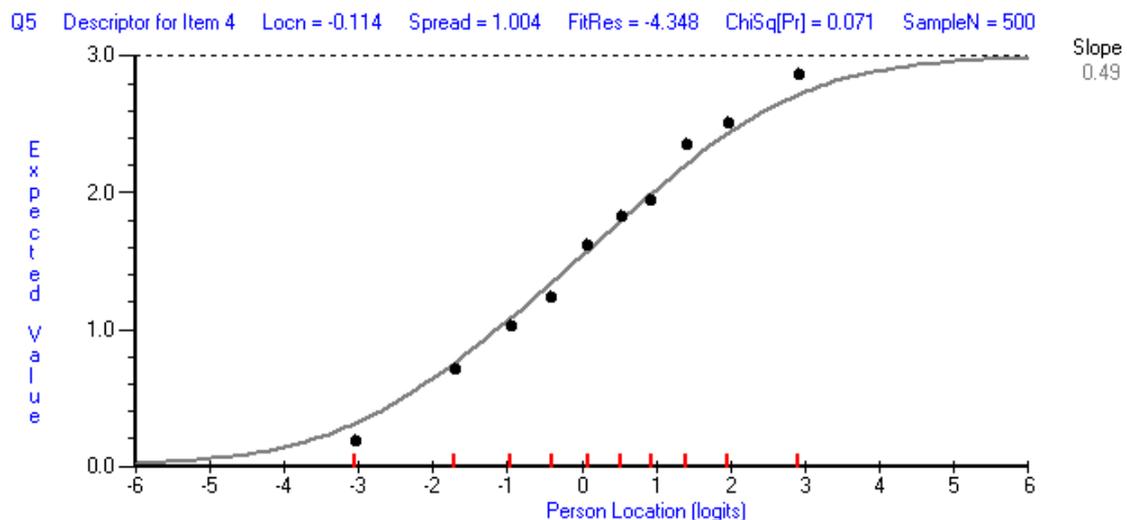


Figure 19 ICC of Q5 (*Originality of this production is....*)

Comment: The assessment data (dots) for this criteria item also tends to over-discriminate a little across total score groups.

## Q6 CLARITY

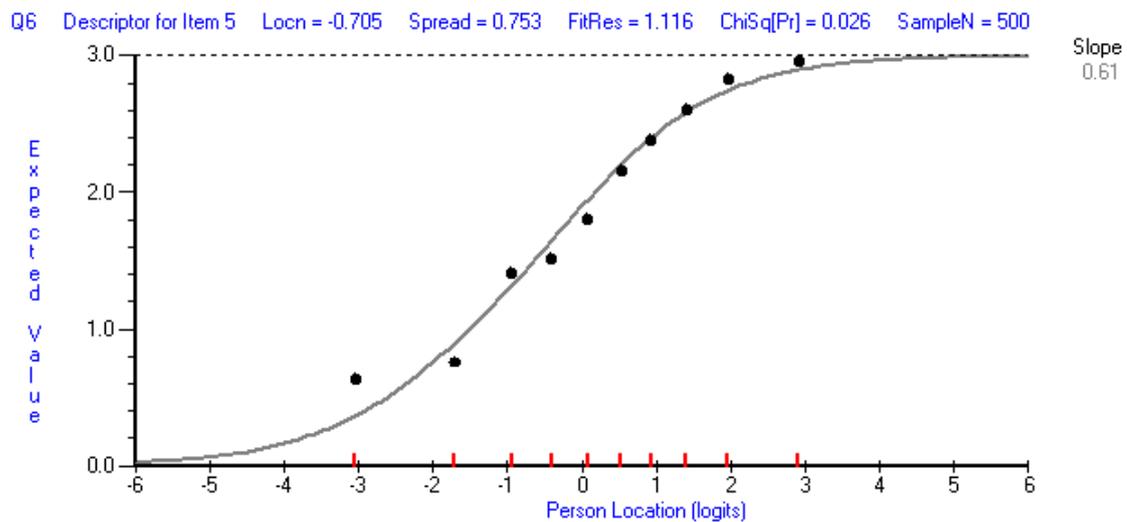


Figure 20 ICC of Q6 (*The clarity of the theme....*)

Comment: The assessment data for this criteria item (dots) shows some under-discrimination across the lowest score group.

## Q7 ARTISTIC QUALITY

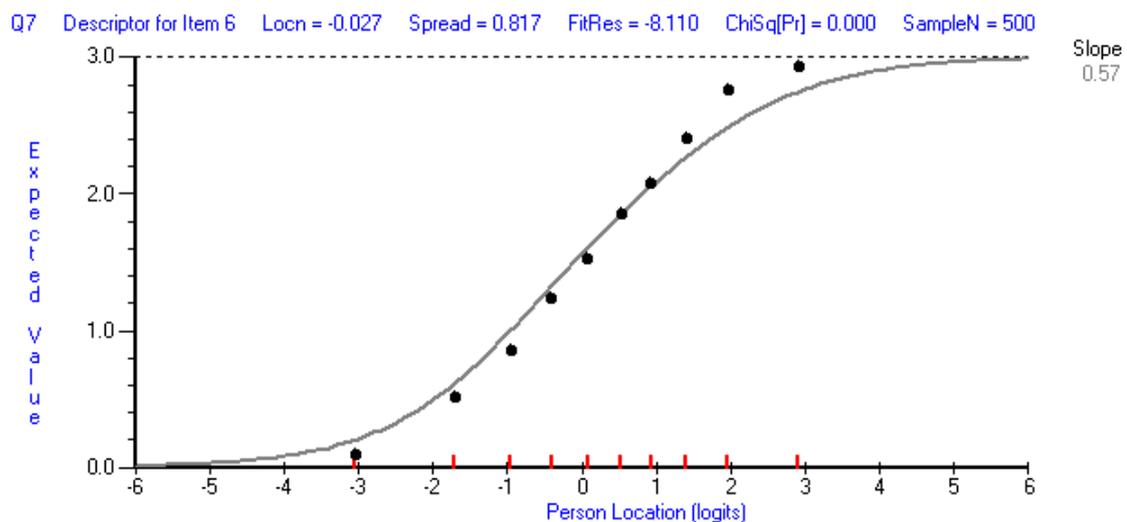


Figure 21 ICC of Q7 (*The artistic quality of this production is....*)

Comment: The assessment data (dots) for this criteria item tends to over-discriminate a little across total score groups.

## Q8 ENHANCEMENT

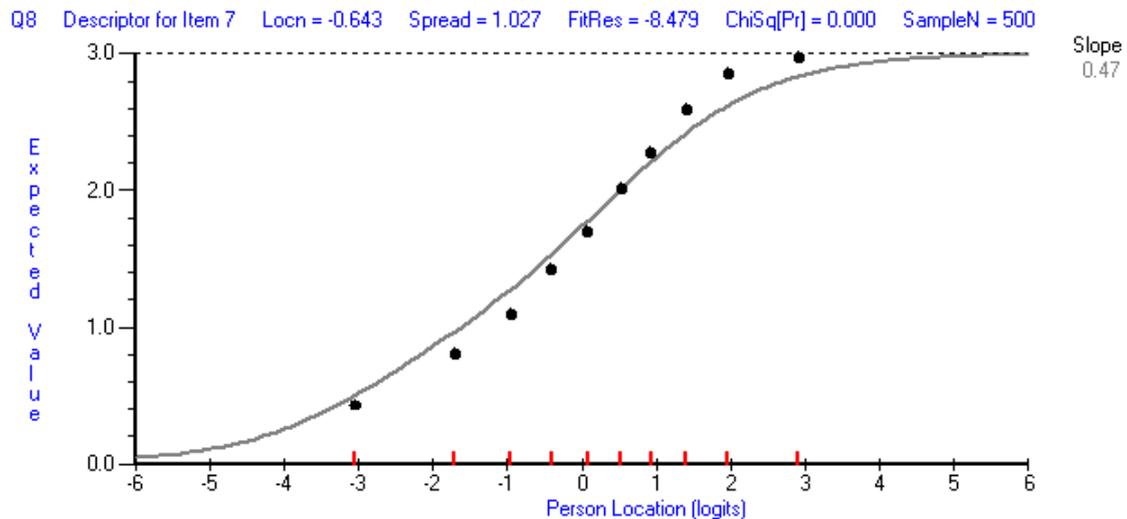


Figure 22 ICC of Q8 (*Different elements enhance each other*)

Comment: The assessment data (dots) for this criteria item tends to over discriminate a little across total score groups.

## Q9 INSPIRATION

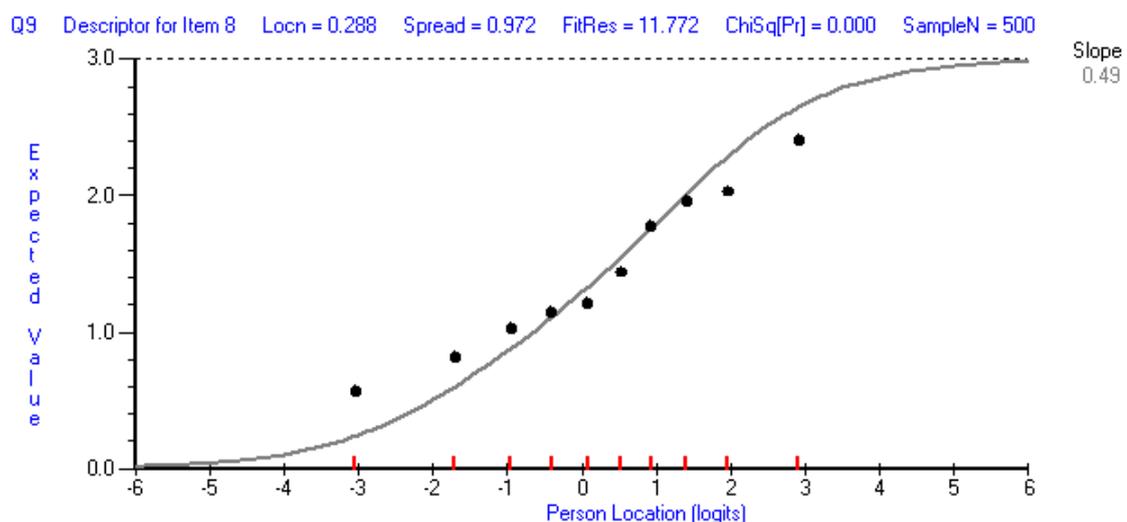


Figure 23 ICC of Q9 (*This production inspires thoughts of other narratives...*)

Comment: The assessment data (dots) for this criteria item tends to under-discriminate a little across total score groups.

## Q11 GESTALT

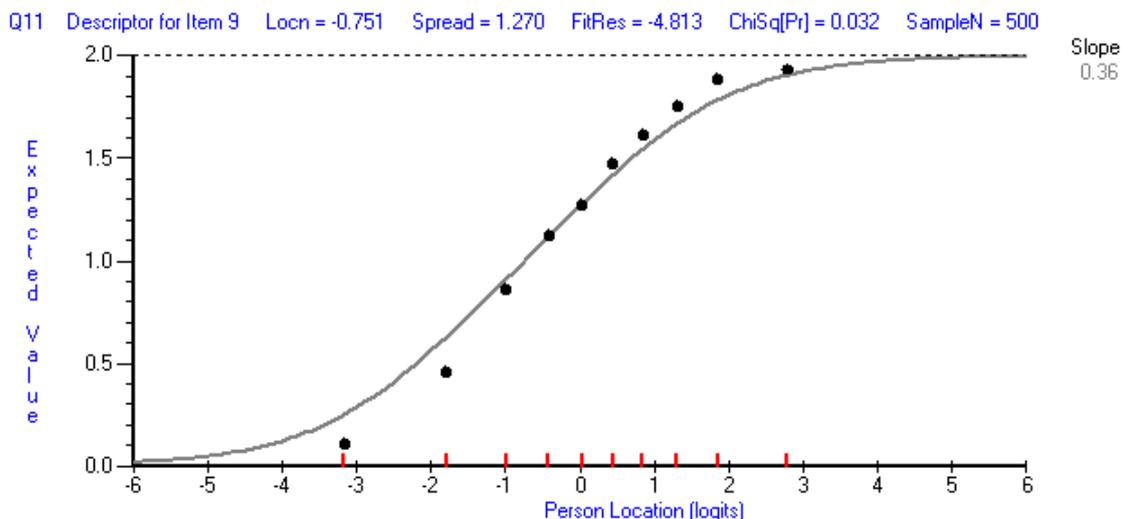


Figure 24 ICC of Q11 (*The best element of this production is....*)

Comment: The assessment data (dots) for this criteria item shows some over-discrimination across total score groups.

## Q12 INTELLECTUAL LEVEL

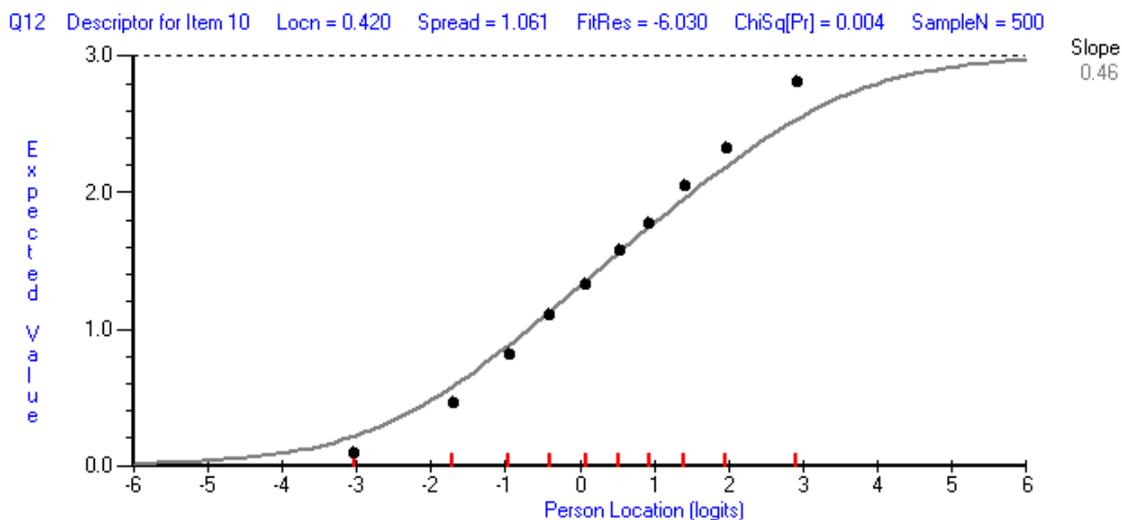


Figure 25 ICC of Q12 (*The intellectual level of this production is....*)

Comment: The assessment data (dots) for this criteria item shows some over-discrimination cross total score groups.

## Q13 TIMING

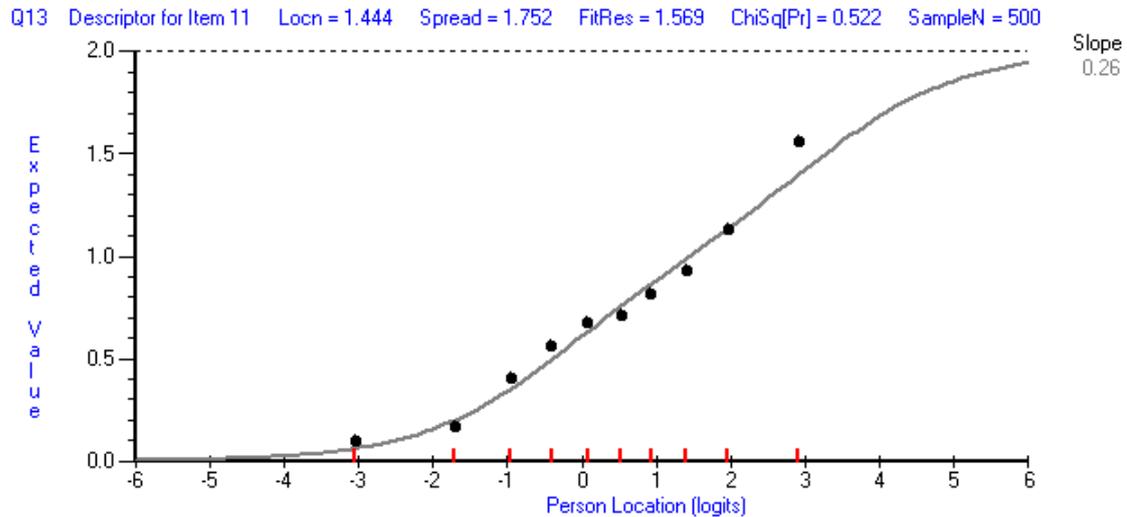


Figure 26 ICC of Q13 (*The duration of this production seems...*)

Comment: The assessment data (dots) for this criteria item shows very little over-discrimination except for the highest total score group which tends to score a little higher than expected.

## Q14 PRODUCTION QUALITY

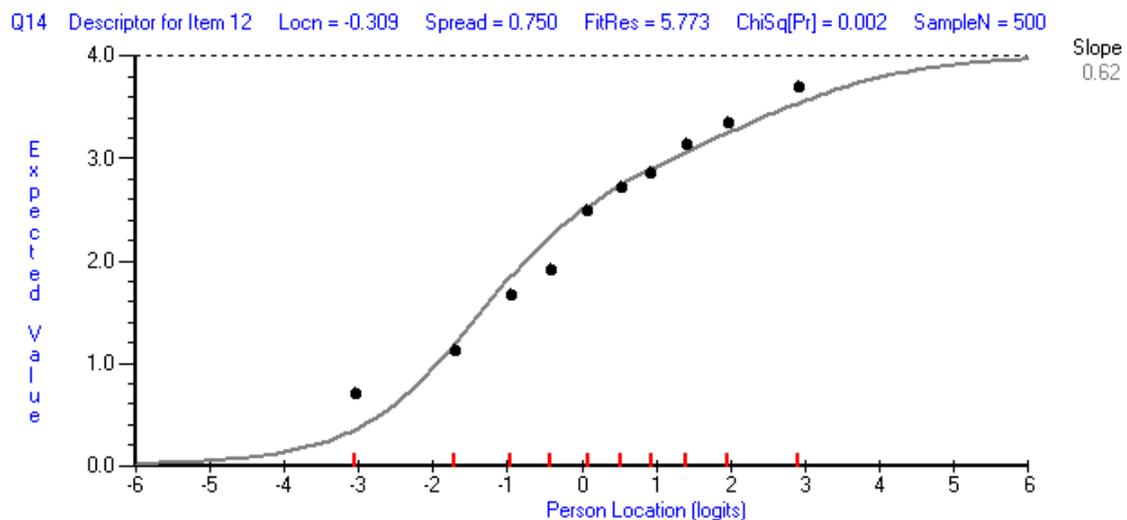


Figure 27 ICC of Q14 (*Overall this production is Imitative (0), Superficial (1), Ordinary (2), Thoughtful (3), Insightful (4)*)

Comment: The assessment data (dots) for this criteria item shows little under-discrimination across lower score groups and very little over-discrimination in highest score group.

## Q15 EMOTIVENESS

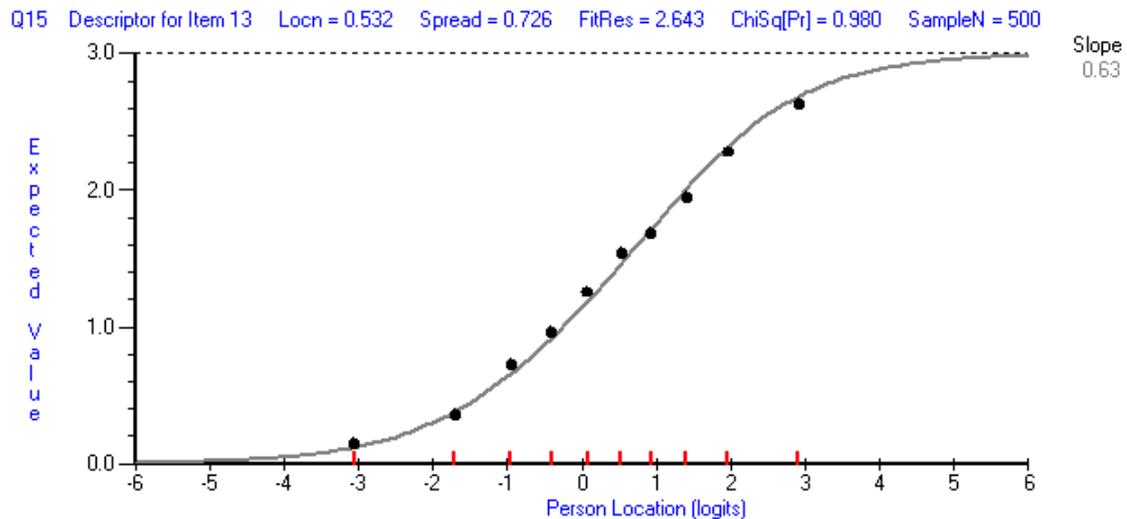


Figure 28 ICC of Q15 (*Emotive level of this production is ....*)

Comment: The assessment data (dots) for this criteria item shows a very nice – excellent fit to the ideal curve (line).

## Q16 BELIEVABILITY

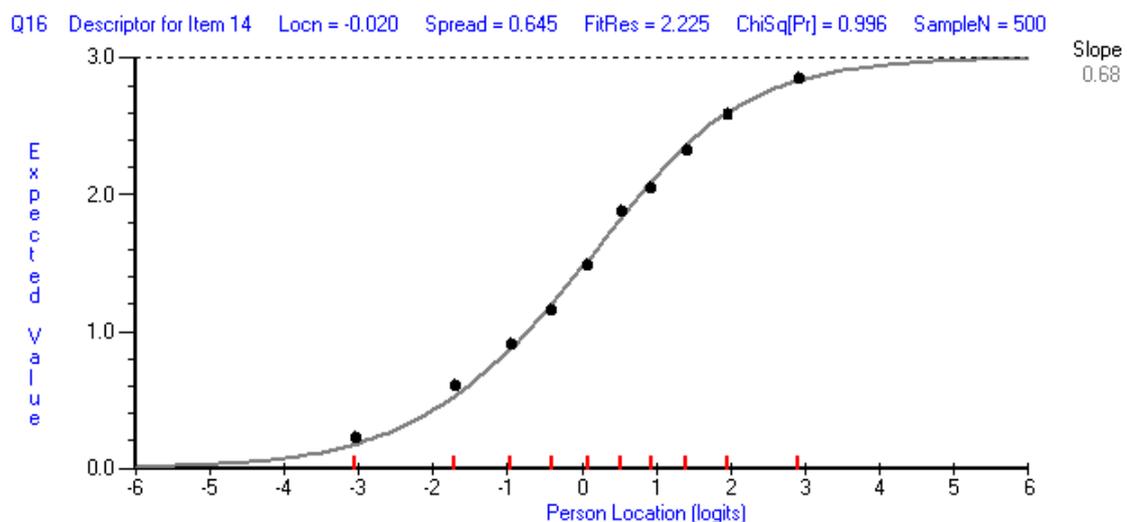


Figure 29 ICC of Q16 (*Believability of this production is.*)

Comment: The assessment data (dots) for this criteria item also shows a very nice – excellent fit to the ideal curve (line).

## Q17 SOCIAL RELEVANCE

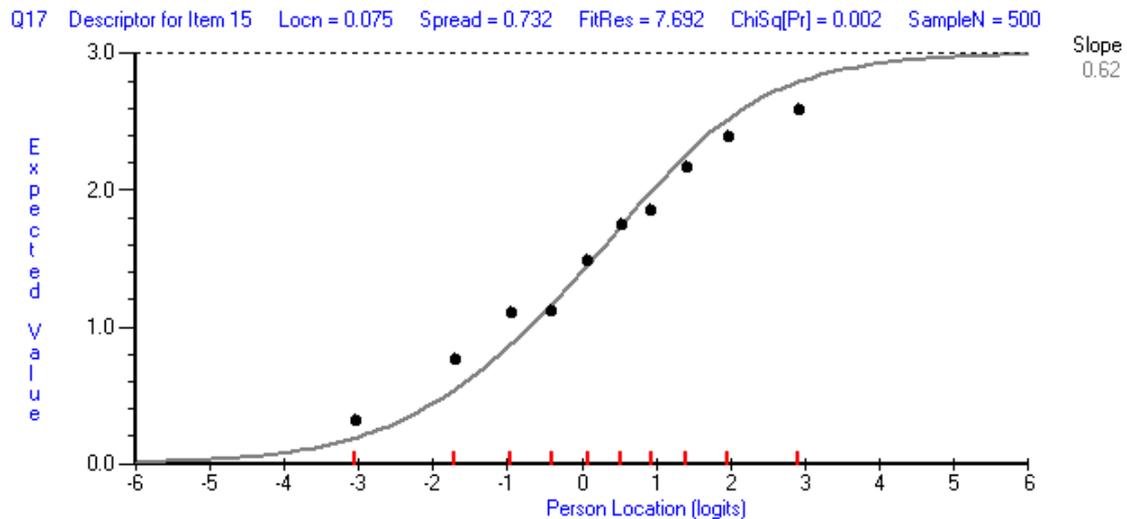


Figure 30 ICC of Q17 (*Social relevance of this production is...*)

Comment: The assessment data (dots) for this criteria item shows tendency to under-discriminate across total score groups.

## Q18 PRODUCTION VALUE

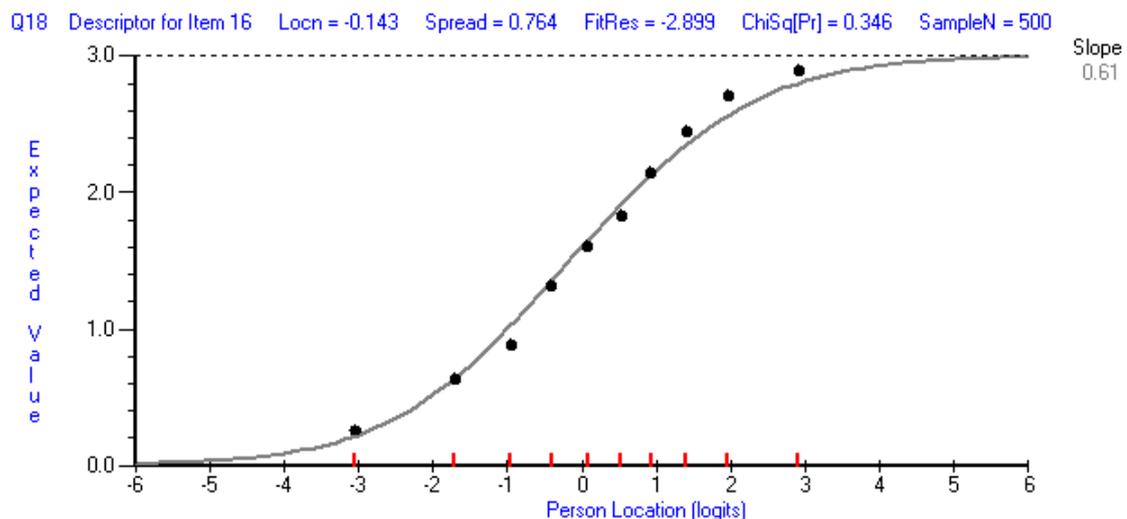


Figure 31 ICC of Q18 (*Overall production values are...*)

Comment: The assessment data (dots) for this criteria item shows a very good fit to the ideal curve (line) – slight over-discrimination in highest scoring groups.

## Q20 ETHICS

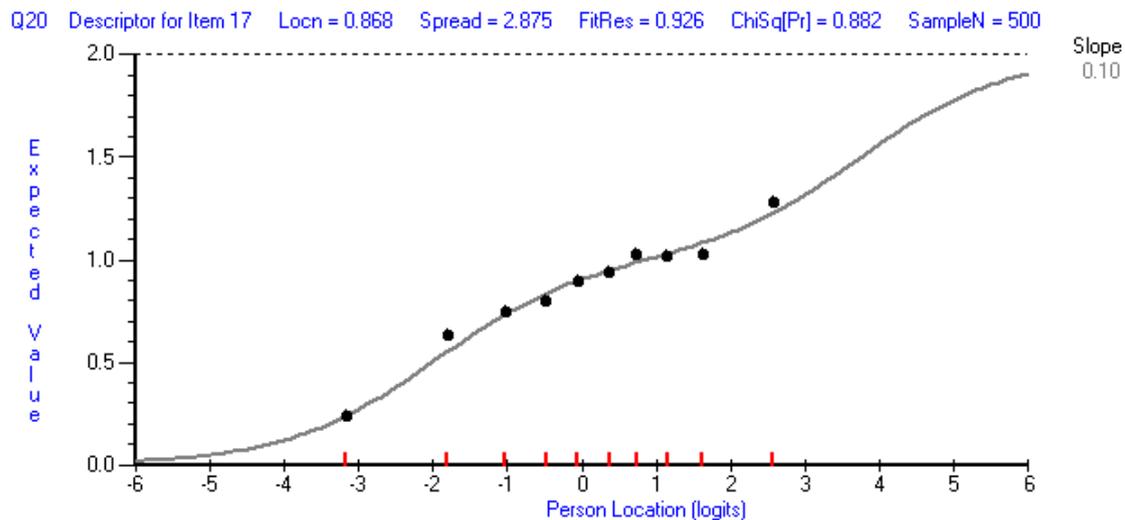


Figure 32 ICC of Q20 (*The use of emotions in this production is...*)

Comment: The assessment data (dots) for this criteria item shows a very good fit to the ideal curve (line). Wobbliness of expected curve is due to placement of the criteria scale thresholds locations (see Figure 17 above).

## Q21 INNOVATION

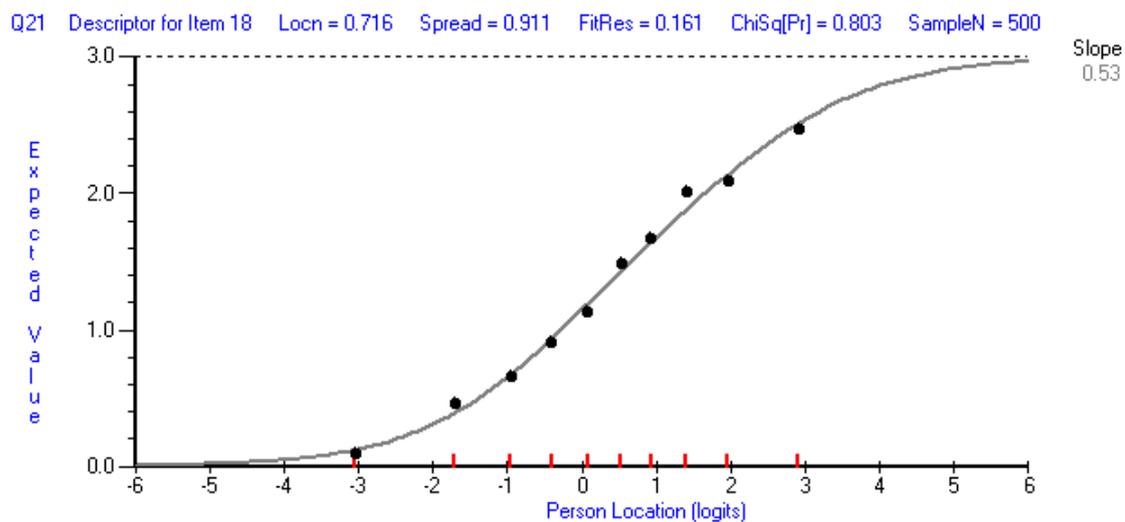


Figure 33 ICC of Q21 (*The level of innovation in this production is...*)

Comment: The assessment data (dots) for this criteria item shows a very good fit to the ideal curve (line).

## Q22 AIMS

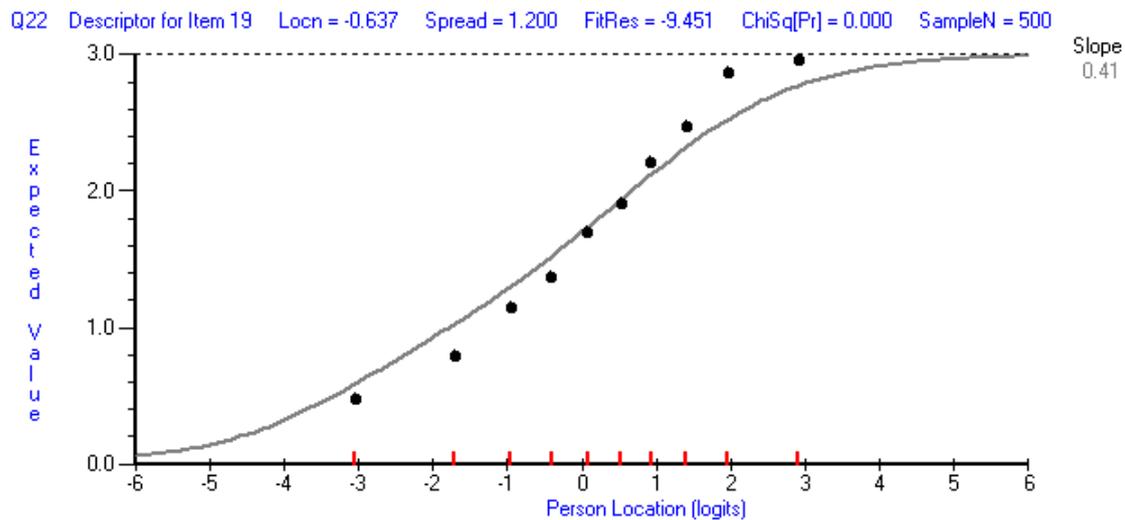


Figure 34 ICC of Q22 (*This production has attained its projected aims*)

Comment: The assessment data (dots) for this criteria item tends to over-discriminate across total score groups (lower total score groups tending to score less than expected and higher total score groups tending to score higher than expected).

## Q23 WORK EVIDENT

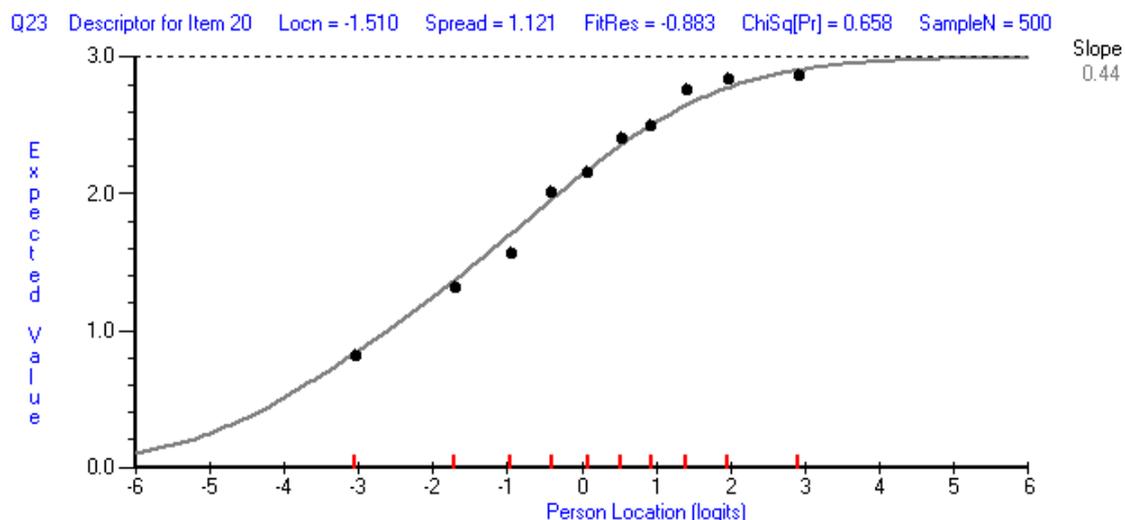


Figure 35 ICC of Q23 (*The amount of work evident in this production is...*)

Comment: The assessment data (dots) for this criteria item shows a pretty good fit to the ideal curve (line).

## Q24 EXHIBITION

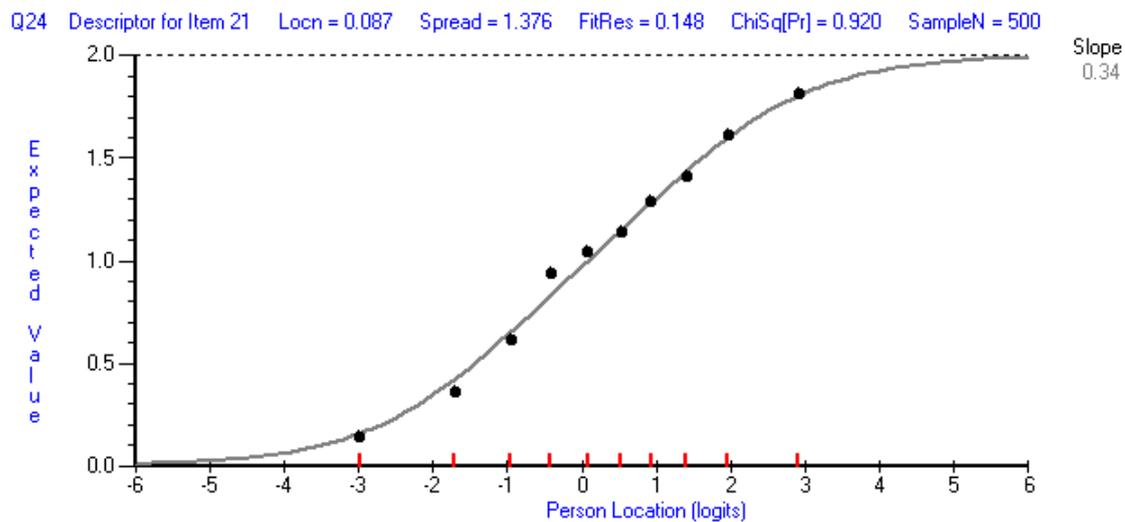


Figure 36 ICC of Q24 (*Your recommended exhibition site for this production is...*)

Comment: The assessment data (dots) for this criteria item also shows a pretty good fit to the ideal curve (line).

## 4. Conclusion

The aim of the project was to show that assessment of image-based creative works is as consistent as assessment conducted in traditional discipline areas. Thirty (30) screen production academics from 22 institutions were tested using the same sample of 45 short honours productions and a carefully selected set of 34 criteria. Creative works alone were assessed without any written components or exegeses. Rasch psychometric modelling was then used to test the internal consistency of the assessment data against an ideal sample comparison.

The results confirmed the hypothesis that screen production assessors are consistent even when compared to an ideal assessment outcome. Accordingly, the SPA scale can be accepted as a valid and highly reliable measure of the quality of screen productions at tertiary level. The assessment criteria, though covering a range of aspects, are accepted as being internally consistent and well-targeted to this particular sample of productions.

There was also little difference in the assessments of “objective” and “subjective” criteria which invites all kind of questions regarding the status of subjective and objective verification procedures within creative arts.

## Notes and references

1. *The Birth of a Nation* directed by D.W. Griffith in 1915 was 190 minutes long and is generally considered to be the first “feature” film.
2. Weeden, W.J. (Committee on Educational Television Services) *Report of the Advisory Committee on Educational Television Services to the Australian Broadcasting Control Board*, Canberra, Commonwealth Govt. Printer, 1966.
3. Paterson, Barbara, *Renegades: Australia’s First Film School from Swinburne to VCA*, The Helicon Press, 1996, p. 10.
4. For details see Australian Film Television and Radio School, *The AFTRS book: a story about change*, AFTRS, 2010.
5. ARC, 2008 Funding rules, 6.5.1b, [http://www.arc.gov.au/pdf/DP09\\_FundingRules.pdf](http://www.arc.gov.au/pdf/DP09_FundingRules.pdf), accessed November 2010.
6. Much of this development was based on the writing of Ferdinand Saussure and writers such as: Roland Barthes, Vladimir Propp, Levi-Strauss, Fisk and Heartly, Antony Wilden, Christina Metz to name only a few. For more details see: Saussure, Ferdinand, *Course in General Linguistics*, Glasgow, Fontana/ Collins, 1974; Barthes, Roland. *Image-Music-Text*, Fontana/ Collins, 1977; Propp, Vladimir. *Morphology of the Folk Tale*, trans. Laurence Scott, Austin & London, University of Texas Press, 1958; Levi-Straus, Claud. *Structural Anthropology*, Harmondsworth, Penguin Books, 1972; Fiske, John & John Hartley. *Reading Television*. Routledge, New York. Publication Year, 2003; Wilden, Antony. *System and Structure*, London, Tavistock, 1972; Metz, Christian. *Language and Cinema*, trans. Donna Jean Umiker-Sebeok, The Hague, Mouton, 1974.
7. Australian Department of Employment, Education and Training, ‘Dawkins White Paper’, <http://catalogue.nla.gov.au/Record/455837>, accessed November 2010.
8. Ross, David, *Aristotle The Nicomachean Ethics: Translated with an Introduction*. Oxford, Oxford University Press, 1925.
9. Rasch, G. (1960/80). *Probabilistic models for some intelligence and attainment tests*, (Copenhagen, Danish Institute for Educational Research). Expanded edition (1980) with foreword and afterword by B.D. Wright, (1980). Chicago: The University of Chicago Press.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.). *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. IV* (pp.321-334). Berkeley CA: University of California Press.

Professor David Andrich, who is internationally known in the field of statistical measurement and for Rasch psychometric modelling in particular, was the primary source of statistical advice for this project. In this context see his:

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43 (4), 561-574.

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological Methodology*, San Francisco, Jossey-Bass. (Chapter 2, pp. 33-80.).

Andrich, D. (1988). *Rasch models for measurement*. California: Sage Publications.

Andrich, D. & Styles, I. (2004). *Report on the Australian Early Development Inventory*. Perth, Western Australia: Institute for Research in Child Health

Andrich, D. (2010) Educational Measurement: Rasch Models. In *International Encyclopedia of Education* 3rd Edition. Edited by Eva Baker, Penelope Peterson and Barry McGaw, Elsevier.

10. For definitions of Gestalt see D. Brett King and Michael Wertheimer, *Max Wertheimer and Gestalt Theory*. New Brunswick NJ, Transaction Publishers, 2005.
11. Pearson Psychometrics Laboratory (UWA) can be found at the following web address: <http://www.education.uwa.edu.au/ppl>
12. Andrich, D., Sheridan, B. & Luo, G. (2010). *Rasch models for measurement: RUMM2030*. RUMM Laboratory, Perth, Western Australia.
13. Most of the results in this section were commissioned for this project from the Pearson Psychometrics Laboratory (UWA). Much of the report was prepared by Dr Irene Styles, Senior Research Fellow with the Pearson Psychometrics Laboratory.
14. Figure 3 shows locations (in logit units) against raw total scores for the scale. It may be seen that a location of 2.00 logits corresponds to a raw score on the SPAS of about 55. A location of -1.00 logits corresponds to a raw score of about 22. Thus the mean location (0.095) for all the locations in Figure 18 corresponds to a raw score of about 35.  
Most cinematographers will recognize the above curve as being very similar to the characteristic curve that represents the film's sensitivity. Indeed, the curve could well be a plot of the amount of exposure (x-axis) against the grain density achieved by that exposure (y-axis).  
[http://motion.kodak.com/motion/uploadedFiles/US\\_plugins\\_acrobat\\_en\\_motion\\_newsletters\\_filmEss\\_06\\_Characteristics\\_of\\_Film.pdf](http://motion.kodak.com/motion/uploadedFiles/US_plugins_acrobat_en_motion_newsletters_filmEss_06_Characteristics_of_Film.pdf)
15. Only one of the 45 productions in the assessment sample was based on Computer Generated Effects (CGE) and there was only one animation. Accordingly, the characteristic curves for CGF and Animation were not included in this collection of diagrams as the range of the results was statistically small. Also, the art design characteristic curve was not available for this publication.

## Appendix 1

The *Assessing of Screen Production Outputs in Nineteen Australian Film Schools* was an Australian Learning and Teaching Council (ALTC) funded initiative conducted as a partnership between the National Academy of Screen and Sound (NASS) Research Centre, Murdoch University, and Victorian College of the Arts (VCA), The University of Melbourne, Sydney University of Technology (UTS), Royal Melbourne Institute of Technology (RMIT) University, Griffith University, Flinders University and in association with ASPERA.

### Priority Project Grant Scheme

This study was funded in 2008 under the Australian Learning and Teaching Council Priority Projects Program, and the priority area of academic standards, assessment practices and reporting.

### Project Team

The project team covered five Australian states (including Australian Capital Territory) with each collaborator playing an active role in implementing the test and final assessments in their state. Team members collaborated over the two conferences and other dissemination activities. The team built upon networks established through ASPERA and consisted of the project leader Dr Josko Petkovic and five other collaborators:

Professor Ian Lang, VCA, The University of Melbourne  
 Mr Leo Berkley, RMIT University  
 Associate Professor Gillian Leahy, University of Technology, Sydney  
 Mr Nicholas Oughton, Griffith University;  
 Ms Alison Wotherspoon, Flinders University.

Project Manager:

Ms Linda Butcher, Educational Development Unit, Murdoch University.

External Assessor:

Professor Su Baker, Director of the VCA, The University of Melbourne.

### Statistical Analysis team:

Professor David Andrich, Director Pearson Psychometrics Laboratory at the University of Western Australia; and  
 Dr Irene Styles, Principal Research Fellow for Pearson Psychometrics Laboratory at the University of Western Australia.

### State Coordinators and Assessors

#### WEST AUSTRALIA

Dr Josko Petkovic, Murdoch University (WA Coordinator)  
 Dr George Karpathakis, Edith Cowan University  
 Dr Ken Miller, Curtin University  
 Dr Larissa Sexton Finck, The University of Western Australia  
 Associate Professor Martin Mhando, Murdoch University

**VICTORIA**

Professor Ian Lang, The University of Melbourne (Victorian Co-Coordinator)

Leo Berkeley, RMIT University (Victorian Co-Coordinator)

Nicolette Freeman, VCA, The University of Melbourne

Jill Holt, Swinburne University

Associate Professor Leon Marvell, Deakin University

Christine Rodgers , RMIT University

**NSW**

Associate Professor Gillian Leahy, University of Technology, Sydney (NSW Coordinator)

Professor Ross Harley, College of Fine Arts, The University of New South Wales

Associate Professor Hart Cohen, University of Western Sydney

Dr Maree Delofski, Macquarie University

Tim Thomas, University of Canberra

**QUEENSLAND**

Nicholas Oughton, Griffith University (Queensland Coordinator)

Helen Yeates, Queensland University of Technology

Geoff Portmann, Queensland University of Technology

Charles Strachan, Griffith University

Associate Professor Michael Sergi, Bond University

**SOUTH AUSTRALIA**

Alison Wotherspoon, Flinders University (SA Coordinator)

Shane McNeil, ex Flinders University

Cole Larsen, Flinders University

John McConchie, Flinders University

Ian Dinning, University of South Australia

Helen Carter (Trial-run assessor)

Ian Hutchinson (Trial-run assessor)

**UNITED KINGDOM**

Dr Zemirah Moffat, University of Kent (UK Co-Coordinator)

Dr Tony Downton, Goldsmiths College, University of London (UK Co-Coordinator)

Prof Joram ten Brink, University of Westminster

Dr Alison Kahn, The Oxford Academy of Documentary Film (OADF)

Dr Rachel Garfield, Goldsmiths College, University of London