Investigating the Potential for RGT and ACJ towards deeper insights of Teacher Assessment Practices


Eva Hartell, KTH Royal Institute of Technology, ehartell@kth.se
Helena Isaksson Persson, KTH Royal Institute of Technology, helenaip@kth.se
Scott Bartholomew, Purdue University, sbartho@purdue.edu
Greg Strimel, Purdue University, gstrimel@purdue.edu

Abstract

The evolution of advanced technology systems and the labour market for future engineers and designers are a global matter. In light of this reasoning, a global perspective on technology education becomes even more important. Assessment is key in order to bridge teaching and learning and an international perspective is needed for understanding different assessment practices in technology education. The purpose of this paper is to investigate potential methods of gaining new perspectives and understanding of teacher assessment practices. Adaptive comparative judgement (ACJ) is an assessment method that has been proven to provide valid, reliable, and feasible results for the assessment of open-ended design problems within technology/engineering education in several countries (Hartell & Skogh, 2015; Kimbell, 2012; Power & Seery, 2012; Seery, Canty, & Phelan, 2011; Bartholomew, 2016). ACJ has also been used as an approach to compare teachers’ assessment practices across countries (see e.g. Bartholomew et al, 2017). Reparatory grid theory (RGT) is a method based on George Kelly’s theory of personal constructs (Kelly, 1963). RGT is used to explore informants’ interpretations and views, on certain topics, for example products or other artefacts (Isaksson Persson, 2015), and teachers’ assessments of portfolios in crafts/sloyd and technology education (Björklund, 2008; Lindström, 2001). The results of ACJ for assessment can be represented in a quantitative manner (Pollitt, 2012) and can be complemented with qualitative measures of think-aloud protocols and or comments from informants during the judgement sessions (see e.g. Hartell & Skogh, 2015).

This paper will explore the potential for, and implications of, combining RGT and ACJ outputs a richer understanding of teachers’ assessment values when assessing open-ended students design portfolios and products by deploying RGT on think-aloud protocols and comments provided by judges during ACJ.

Key Words: Adaptive comparative judgement, comparative judgement, reparatory grid theory, assessment, technology education, engineering education, STEM education
1. INTRODUCTION

Establishing an understanding of teachers’ assessment practices with regards to design is paramount and the need to analyse teachers’ values in depth is both needed and asked for (Bjurulf, 2011; Hartell, 2015; Williams, 2016). This paper reports the results from a pilot study that was undertaken with the purpose of examining the potential (if any) of using the Reparatory Grid Technique (RGT) (Kelly, 1963) as a tool for analysing the feedback comments, collected from teachers, as they assess student design work through an Adaptive Comparative Judgement software engine titled CompareAssess. This pilot study applies RGT on data collected through an on-going project investigating teachers’ assessment practices and values in open-ended design across cultures in USA, UK, and Sweden (see Bartholomew, Hartell, & Strimel, 2017). We investigated the potential of RGT as a tool for deeper analyses of judge comments the feedback comments collected from Swedish teacher judges as they compared student’s design work (design products and portfolios). The results highlight the potential utility of the RGT for analysing the qualitative data collected through ACJ assessment software.

1.1 Reparatory Grid Theory

The Reparatory Grid Technique (RGT) is derived from George Kelly’s work and is based on his Theory of personal constructs (Fransella, Bell & Bannister, 2004; Jankowicz, 2004; Kelly, 1963). According to this theory, people interpret and perceive the world based on their previous experiences and, as people gain new experiences, their perception of the world changes. When people interpret the world, they use multidimensional attributes, which Kelly calls constructs (Kelly, 1963). Fransella, Bell, and Bannister (2004) summarise Kelly’s view on how people construe the world as “[…] we never affirm anything without simultaneously denying something” (Fransella, Bell, & Bannister, 2004, p. 7). Correspondingly, we hypothesize that a deeper understanding may be potentially obtained by examining their constructs, as obtained through an ACJ judges’ comments, using the RGT.

A construct has a bipolar structure. To fully understand a person’s view on a topic we need to know both dimensions of a construct. The informant can easily express one dimension, in the analysis called the emergent pole. Contrasting to this pole is the implicit pole (Fransella, Bell, & Bannister, 2004; Jankowicz, 2004). The characteristic of the implicit pole is usually defined as a contrasting quality of the emergent pole. The procedure of the RGT results in several two-dimensional constructs through the production of both implicit and emergent poles.

Every individual holds a complex and unique set of constructs. However, Kelly (1963) argues that people can belong to a cultural group that may construe experiences in the same manner as other members of that group. In this study, we hypothesize that the Swedish technology teachers, who acted as judges using ACJ in this project, all belong to a group of individuals with similar experiences concerning assessment of design and technology. For example, this group needs knowledge of the design process and should possess similar abilities, and experiences, in the analysis of the technological world. As artefacts are results of the design process, this study is inspired by RGT studies where products and other artefacts are used as sources for elicitation of constructs. A few previous examples include: consumer products (Persson, Hiort af Ornäs, & Jordan, 2007), fine metal craft (Lindström, 2001), artefacts made by pupils during lessons in technology (Björklund, 2008), consumer products (Isaksson Persson, 2015), and teaching materials (Isaksson Persson, & Gumaelius, 2016).

To elicit a construct using RGT, a common method is to present a triad of elements representing a certain topic (Fransella, Bell, & Bannister, 2004; Jankowicz, 2004). The informant then must choose two elements that have shared characteristics, which separate these items from the third element. The characteristic(s) of the two elements is the emergent pole of a construct and the characteristic(s) of the third element are the implicit pole. There are several variations of this method, for example one can use only two elements where the informant discusses how the elements are alike or different. If they are different, the two poles are specified. If they are alike the informant is asked to describe the opposite of the similarity (Fransella, Bell, & Bannister, 2004). For more detail on RGT procedure see (Isaksson Persson, 2016).

In this study, the judges’ ACJ assessment procedure was compared to a RGT procedure, were two elements are compared and similar/dissimilar, traits are identified. Further, one common procedure in RGT studies is that of
comparing and rating all elements to find coherence among constructs and elements. While this secondary procedure was not performed in this study we argue that through the ranking procedure of ACJ coherence between the elements may be established.

1.2 Adaptive Comparative Judgement (ACJ)

ACJ is an assessment method where judges (e.g. teachers) are presented with two pieces of student work (e.g. design portfolios), which they compare in terms of quality. The judges may base their decisions on their professional expertise; some predetermined criteria, and/or their personal beliefs. Rather than assigning any sort of score, judges are simply asked to choose one item of student work over another when presented with a pair. Following each decision, the judges are presented with another pair and the process repeats itself, producing a rank order, or rather a continuum of student work, which signifies quality. This process repeats itself until a pre-determined time or reliability level (i.e., $r > .9$) (Pollitt, 2012). Additional statistics for further analysis are produced (e.g. Rasch model misfit statistics, judges’ time in judgments). The ACJ process, and accompanying software programs (e.g., CompareAssess) which facilitate the judgments, can also allow judges to provide motives for their decisions which include the rationale on which they based their judgment (Bartholomew et al 2017; Hartell & Skogh, 2015). For a more in depth explanation of ACJ, we direct the reader to Pollitt (2012).

Within technology and engineering education, several studies have examined both the validity and reliability of the ACJ assessment method (Bartholomew, Hartell, & Strimel, 2017; Bartholomew & Yoshikawa, 2018; Hartell & Skogh, 2015; Kimbell, 2012; Seery, Canty, & Phelan, 2011). These studies have all shown high levels of consistency in teachers’ judgement, even on open-ended tasks. Relatedly, Hartell and Skogh (2015) tried to unpack what teachers value as criteria for success, while assessing multimodal portfolios, having teachers provide a motive for choosing one item as better by thinking aloud during the ACJ process, which was recorded by a MP3 recorder, and then transcribed into protocols. These think aloud protocols were analysed and the study found that teachers value how students present the narrative within their portfolio. In Bartholomew et al (2017) judges from USA, UK/Ireland, and Sweden were asked to assess an open-ended design task using ACJ and provide comments in terms of feedback used for their comparative judgements. These comments have been analysed showing both consistency, and differences, between the groups of judges in various locations and cultures (see Bartholomew et al, 2018).

1.3 Reparatory Grid Technique and Adaptive Comparative Judgement

With the purpose of a deeper understanding into teachers’ assessment practices and further insight into what teacher’s emphasise when assessing students’ work, this pilot study investigated a combinational effort of RGT and ACJ. In this investigation, the judges’ ACJ assessment procedure was looked upon as similar to an RGT procedure were two elements are compared and traits are identified (both similar and different). Further, one common procedure in RGT studies is that of comparing and grading all elements to find coherence among constructs and elements. While this secondary procedure was not performed in this study, we argue that the ranking procedure of ACJ can establish similar coherence between the elements.

2. METHODOLOGY

Data from an on-going research project investigating teachers’ assessment practices and their values during ACJ process was used in this study to explore RGT as tool for the analysis of the judge comments collected via the ACJ process. A member of the research team collected design products and portfolios from American middle school students (12–14 year old, $N = 760$), who designed pill dispensers for elderly people who enjoyed travelling (see Bartholomew, Reeve, Veon, Goodridge, Stewardson, Lee, & Nadelson, 2017). The students in this study worked in groups and produced products ($N = 175$) and process portfolios ($N = 175$), which were then collected for ACJ assessment by three groups of judges to assess the student work (Bartholomew et al, 2017). Judges were recruited from the USA, UK/Ireland, and Sweden and were trained in the ACJ process for this task by the researchers involved. The judge comments collected from four Swedish judges were examined for the purposes of this pilot effort. These judges all had a background in technology education in Sweden and two of them were STEM teachers in teaching years 1–6 and 4–9, one was a technology teacher for years 7–9, and one was a professor of technology education.
The analysed ACJ data included a rank order of the design work from the American middle school students and feedback comments from the Swedish judges that were collected during the ACJ process when they judged the student work. The RGT analysis includes 175 product pictures and 459 judge comments. Further analysis focused on student work ranking from 1 – 5 from the top (called W1, W2, W3, W4, and W5). The rationale for selecting these pieces of student work was to get a reasonable amount of data to achieve the overall goal for this pilot, which is to investigate the possibilities for extensive analysis of ACJ data with RGT. As a start, products with high rank order were selected as they may have potential to reflect qualities that are considered as important to the judges.

3. DATA ANALYSIS AND RESULTS

ACJ has similarities with construct elicitation in the RGT. A judge’s comment defines a quality/characteristic that describes a *difference* between two elements. The difference between two elements is described in a written comment by the judge. This description of a difference may therefore be similar to the descriptions of contrasting characteristics representing the emergent and implicit pole in a construct elicited with RGT.

Table 1 provides an example of how ACJ data could potentially be arranged in alignment with RGT. To the left is product Work 1 (W1), which received a ranking as the best product overall according to the Swedish judges. All comments this product received from the Swedish judges are presented (Table 1, comments 1-5). Under the heading ELEMENTS (products) are the products that in comparison with W1 elicited the comments from the judges.

In this analysis, comments 4 and 5 are considered to be the carriers of more information than comments 1-3. Comments 4 and 5 are results from assessment sessions where W1 is compared with W2, (rank order two). Comments 1-3 are results from comparisons with W1 and other products not included in top five rankings (rank order 65, 74, 141). In this study all comments are regarded, but a comparison between elements with similarly high rank orders provides an interesting glimpse into the specific qualities appreciated by the judges. In RGT the whole construct, both the emergent and the implicit poles, informs us about the judges’ view on the topic. Construct D and E (Table 1) showcases the comments and product picture of W1 and W2. Based on the comments provided we hypothesize that an important quality to judges was *functionality*, as demonstrated in the judge comments.
It is also interesting to regard the other comments related to W2 (Table 2). In the assessment sessions represented W2 was the winner. In construct F (Table 2) W2 was compared with W4 (rank order four). The winning quality derived from the analysis showed that W2 was better organised than W4 (Table 2, comment 6). Considering comment 4 (Table 1, construct D) we derived that W1 won because it provided a better overview than W2. While overview and organised were important aspects of a winning quality to Judge01, we posit a better understanding of the quality Overview-Organised would be understood if the opposed characteristics were known (Kelly, 1963). In other words, both the emergent pole and the implicit would be useful for further analysis.

<table>
<thead>
<tr>
<th>Swedish Judges ELEMENTS (products)</th>
<th>W2</th>
<th>W65</th>
<th>W74</th>
<th>W141</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergent pole</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct A, A is best in its design and information. (Comment 1, Judge01 comparing W1 and Other)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Construct B, Easy to see if the pills are taken. (Comment 2, Judge03 comparing W1 and Other)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Construct C, It (A) looks to be more functional and it is also simple and stylish. (Comment 3, Judge05 comparing W1 and Other)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Construct D, B provides a better overview. (Comment 4, Judge01 comparing W1 and W2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct E, I liked that it was simple to see if I had missed to take the pills. And I like the construction. (Comment 5, Judge03 comparing W1 and W2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

W2 = rank order
Other = Other than top five elements
Table 2, shows the constructs where W2 is the winner

<table>
<thead>
<tr>
<th>Swedish Judges</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELEMENTS (products)</td>
<td></td>
</tr>
<tr>
<td>Emergent pole</td>
<td></td>
</tr>
<tr>
<td><em>Construct F.</em> B is better organised.</td>
<td>W4</td>
</tr>
<tr>
<td>(Comment 6, Judge01 comparing W2 and W1)</td>
<td>X</td>
</tr>
<tr>
<td>Implicit pole</td>
<td></td>
</tr>
<tr>
<td><em>Construct G</em> Good construction/design.</td>
<td>W2</td>
</tr>
<tr>
<td>(Comment 7, Judge03 comparing W2 and Other)</td>
<td></td>
</tr>
<tr>
<td><em>Construct H</em> B is simple but well-functioning. Well worked/thorough model.</td>
<td>W2</td>
</tr>
<tr>
<td>(Comment 8, Judge05 comparing W2 and Other)</td>
<td></td>
</tr>
<tr>
<td>Wn = rank order</td>
<td></td>
</tr>
<tr>
<td>Other = Other than top five elements</td>
<td></td>
</tr>
</tbody>
</table>

4. DISCUSSION

This initial effort was conducted with the purpose of investigating the potential for RGT as a possible tool for qualitatively analysing feedback comments elicited from ACJ process. Our results revealed promise for using RGT in the further analysis of ACJ comment data. We suggest that, based on our analysis, functionality was important to the Swedish judges as was the quality of Overview-Organised.

Full constructs (Table 1, constructs D and E; Table 2, constructs F) with product pictures representing both poles were elicited while full constructs in words were not elicited in this analysis model (Table 2). According to Kelly (1963) data from a constructs’ both poles are aspects of the same underlying phenomenon. Therefore, data from both poles would likely provide a more robust opportunity for in-depth analysis and understanding.

The experiences from this first pilot study opened further investigation possibilities related to the present data, as well as possibilities to improve the combination of ACJ and RGT. For example the results point to:

- Additional possibilities for analysis related to the scope of this pilot study.
- Image analysis may increase an understanding of the constructs.
- Repeated analysis with the comments from the US and UK judges may reveal additional insights.
- Possibilities exist related to using RGT for expanded data analysis in the scope of the on-going international comparative study on cultural design values (Bartholomew et al, 2017).

The results point to possibilities for further development of the combination of ACJ and RGT.

In this study the judges were asked to provide a motive for why they chose one on top of the other. From this study we have learned that it would be even better if they were given instructions that guide the judges to comment on both objects, this with the purpose to reveal full constructs. Judges need to describe why they chose an object and why they do not chose the other one. To gain richer insights on their motives, another possible option to develop the methodology even further could be to use think aloud protocols where judges think-out loud instead of typing their comments (see Hartell and Skogh, 2015). However, having judges think out loud provides more data but not necessarily richer data.

Complement ACJ with face-to-face interview using RGT.
Conduct similar analysis on qualitative data gained from ACJ trials, in technology education and also other subjects with the purpose of understanding teachers’ assessment practices and feed that into learning activities for students.

5. CONCLUSION

Results from this pilot study highlight possible benefits of combining RGT and ACJ as a method for unpacking judges’ assessment practices, and eliciting evidence of their values and beliefs. As this is a pilot study, one should be cautious to conclude, however the importance of asking judges why they choose and also ask why they do not chose object during ACJ session must be considered in future work with this combined method. From judges’ comments results indicate that judges in this research valued functionality in students’ work.

6. REFERENCES


