Manual Versus Online Card Sorts: 
A Cautionary Tale for IS Researchers 

Completed Research

James S. Denford  
Royal Military College of Canada  
jim.denford@rmc.ca

Kurt B. Schobel  
Royal Military College of Canada  
kurt.schobel@rmc.ca

Abstract

Card sorting is an important and frequently used survey validation technique for information systems (IS) research. A card sort can be time consuming and thus, with the advent of technology, card sorting has been automated with a number of online card sorting programs available for use by researchers. Since the early 1990s, there have been numerous studies that compared online to manual card sorts with most reporting that manual and online card sorts yield similar results. This study found that observation may impact the efficacy between online and manual card sorts thereby casting doubt into the interchangeability of manual and online card sorts. Specifically we find a significant difference between the results of a manual observed card sort and an online unobserved card sort. With the proliferation of card sorts as a construct validation technique, this paper is a caution for IS researchers to be mindful of the design of their cards sort and to be consistent in the delivery of them.

Keywords

Instrument Validation, Survey, Methodology

Introduction

The use of theoretical constructs is a major component of information systems (IS) exploratory research. The process of validating these constructs is equally if not more important than the results obtained. To develop better measures of constructs Churchill (1979) suggested an eight-step process. Steps six and seven called for researchers to assess the reliability of the measures. One means of assessing reliability is through a card sort, a method that started nearly 30 years ago and continues to this day. A card sort is a multi-step process that is used to confirm the coverage of a domain with a set of constructs (Davis et al., 1989, Nahm et al., 2002).

Neufeld (2004) notes the use of card sorts as a data collection technique in the Social Sciences. Additionally, card sorts have been used in usability studies (Faiks & Hyland, 2000; Wood & Wood, 2008; Chapparo, 2008). Nahm et al. (2002) further studied the use of card sorts and found that they are a good measure for assessing the reliability of measurement items and of the validity of the overall construct. A search for “card sort”, “Q-Sort”, or “Item Sort” on ABI/INFORM Global found in excess of 50,000 peer reviewed scholarly papers referring to one of these terms demonstrates the prevalence of card sorting across research domains. Most of these papers discuss card sorts within the methodological component of the paper.

Within the IS discipline, Moore and Benbasat (1991) first operationalized the use of card sorts as a survey instrument validation technique. Since that time, card sorts have been used continuously to validate survey instruments within IS research. Petter (2007, p.639) posited that “a two-step Q-sort may be one of the best methods to assess content validity for formative constructs”. Compeau et al. (2007) used card sorts to validate survey instruments related to the perceived characteristics of innovating (PCI) to extend the work of Moore and Benbasat (1991). In their paper they noted that in 2003 there were 178 citations of the Moore and Benbasat (1991) paper. A search on ABI/INFORMS Global in 2017 puts the number of citations at 468 whereas Google Scholar notes 8015 citations of the paper, indicating that this seminal work is a cornerstone to IS research.
Researchers across all the top IS journals have used card sorts to refine their model constructs before administering surveys (c.f. Bateman et al., 2011; Brown et al., 2016; Komiak & Benbasat, 2006; Larsen, 2003).

Because of the widespread use of card sorts, researchers have looked at ways to improve/simplify the delivery and use of the technique. Consequently, the original manual card sort technique was soon followed by the development of online card sorting tools. The addition of the online card sort brought forth a body of research to test the validity of online card sorts. It is generally felt that there are no significant differences between online and manual card sorts (Rugg et al., 1992; Harper & Van Duyne, 2002; Harper et al., 2003). Our findings indicate that there may be instances when there are differences between the two methods and that IS researchers need to be aware of the implications.

**Methodological Background**

In this section, we review the background of card sorts, their evolution from manual to online applications and the impacts of observation on task performance.

**The Card Sort**

The card sort is also referred to as a Q-Sort or Item Sort. Regardless of the name, when conducting a card sort, a participant is given a set of cards. On each card there is a single statement that represents a possible item within the underlying research construct. The task is to sort the index cards into separate piles or categories based on the similarities and differences among the statements on each card; each pile should reflect one underlying concept of the model. There are two key types of card sorts, open and closed. The difference between a closed and an open card sort is that in the former, the sorting categories are given to the sorter and in the latter, the sorter must define the categories based on the statement on the card and their perceptions of the underlying unifying constructs. The open card sort stage is devised to confirm that the meaning of the construct was well understood, while the convergence and divergence of items within categories is used to demonstrate convergent and discriminant validity (Moore & Benbasat, 1991). As the model and its indicators are refined through the process, it is expected that the reliability will increase from an initial open card sort to a revised closed card sort.

The impetus for this study came from one application of card sorting. The current researchers had conducted an open card sort followed by a closed card sort to validate a survey instrument. The results of the open card sort yielded an acceptable correlation between the research model and the survey questions. We then proceeded to a closed card sort after refining the survey questions based on feedback from the open card sort. The closed card sort yielded a lower correlation between the research model and the survey questions than the open card sort. One would expect the correlation results to improve as a study advances from an open to a closed card sort; yet, this experiment yielded the opposite result. Looking at the conditions of the two experiments we noted that the first card sort was manual and observed whereas the second card sort was online and unobserved which suggested to us that it may be possible that manual and online card sorts may not necessarily be equally effective in all conditions.

**Online Card Sorts**

In the time period since Rugg et al. (1992) completed their study of online versus card sorts, there have been a number of other studies that have supported the substitutability of online and manual card sorts. For example, Harper and Van Duyne (2002) found no significant differences between manual and online card sorts in terms of accuracy, test-retest reliability, and number of categories created.

Much has been said about how to interpret the findings from cards and the methods used to analyze those results. (Moore & Benbasat, 1991; Righi, 2013) and the efficacy of online card sorts (Rugg et al., 1992; Harper et al., 2002; Harper et al., 2003; Paul, 2008). In the first apparent testing of online card sorts Rugg et al. (1992, p.279) found that “No significant differences were found between the types of knowledge elicited by different types of sort”. Their experiment compared a manual card sort to a computerized card sort in the same domain using an open card sort. Harper et al. (2002, p.2049) stated that “a computerized card sorting task is just as effective at eliciting knowledge as a manual card sort”. Harper et al. (2003) reiterate the many benefits of an online card sort and extol the advantages of the
Manual versus Online Card Sorts

software, again using an open card sort. More recently, an experiment was reported wherein the online closed card sort participants performed as well as the ones in the manual sort (Bussolon et al., 2006). Only one study found that computer-based card sorting leads to “a higher number of different possibilities how to assign cards to categories than offline card sorting” and concluded that online and manual closed card sorts may not yield the same results (Greve, 2014, p.42). On the whole, however, evidence points to substitutability of manual and online card sorts and so it was hypothesized that manual card sorts would be completed with the same level of accuracy as online card sort.

H1: Manual card sorts will be completed with the same level of accuracy as Online card sorts.

The Effects of Observation

The debate over whether observation of a subject affects the outcome of an experiment has been hotly contested ever since the results of the Hawthorne Experiments (Roethlisberger & Dickson, 1939) were first published. The notion that an observer has an effect on the people being studied is also referred to as the Hawthorne Effect in reference to the original experiments. Adair (1984, p.334) referred to the Hawthorne Effect as “the problem in field experiments that subjects' knowledge that they are in an experiment modifies their behavior from what it would have been without the knowledge”.

Wickstrom and Bendix (2000) summarized the Hawthorne studies, described the birth of the term “Hawthorne Effect”, outlined the criticisms of the studies and discussed the evaluation of intervention studies. They ultimately stated, “if referring to the Hawthorne studies is thought to add light to the discussion, reference should be done on the basis of a knowledge of the studies and their results” (Wickstrom & Bendix (2000, p.366). This statement is important because there were many concerns over the original findings and the term Hawthorne Effect should not be used arbitrarily.

The medical field regularly tests for the Hawthorne Effect in clinical studies. Most recently, a study was conducted to: determine whether the Hawthorne effect exists, explore under what conditions it might exists, and estimate the size of any effects (McCambridge et al., 2014). Their study concluded that consequences of research participation do exist. Similarly, in an experiment assessing the effectiveness of patient controlled analgesia, it was noted that there is no doubt that the behaviour of research participants is affected by observation (Sedgwick & Greenwood, 2015). Both papers also caution on the generalization of the effects.

With full recognition of the cautions by Wickstrom and Bendix (2000) and in light of the recent studies, a second hypothesis was developed.

H2: Observed card sorts will be completed with greater accuracy than Unobserved card sorts.

Finally, an additional set of hypotheses was created to address each combination of Hypotheses 1 and 2 in order to confirm specific outcomes.

H3a: Online Observed card sorts will be completed with the same level of accuracy as Manual Observed card sorts.

H3b: Online Unobserved card sorts will be completed with the same level of accuracy as Manual Unobserved card sorts.

H3c: Manual Observed card sorts will be completed with greater accuracy than Manual Unobserved card sorts.

H3d: Online Observed card sorts will be completed with greater accuracy than Online Unobserved card sorts.

H3e: Manual Observed card sorts will be completed with greater accuracy than Online Unobserved card sorts.

H3f: Online Observed card sorts will be completed with greater accuracy than Manual Unobserved card sorts.
Experimental Design

The phenomenon of interest was uncovered during the development and validation of a survey instrument, so it was decided a structured approach to confirming the outcome was required; thus, a laboratory experiment was selected as the means to examine the phenomenon of interest. A task-built survey instrument was specifically designed for the experiment to be a highly complex sorting task that would drive participants to relying on the process provided rather than their own cognitive abilities with 79 items in 23 second-order constructs encompassing seven first-order constructs. The instrument was not intended to be applied in a traditional context; rather, it was designed to drive the sorting task and intentionally included closely related constructs to increase complexity. A closed card sort was selected as the number of categories of outcomes could be known a priori and thus success rates could be calculated and compared. An extract from the experimental instrument is included at Appendix 1.

The experimental design was a two-condition random sample design, where student participants (2nd through 4th years) from an undergraduate program in business administration were invited to participate in a closed card sort exercise to assist the researchers in the development of a survey instrument. Upon registering for the exercise, students were assigned to a random group – either manual observed, manual unobserved, online observed, or online unobserved. If they were assigned to one of the two unobserved conditions they were provided a package to be taken away (manual sort) or a hyperlink to complete the card sort (online sort). If they were assigned to one of the observed condition they were individually taken to a room where one of the researchers supervised the card sort. Supervised tasks were undertaken by a single individual at a time, not in a group setting. In order not to give the supervised group an advantage over the unsupervised group, if questioned on an item the researcher declined to answer responding that the interpretation of the items by the participant was what was the card sort sought to understand. Those in the unobserved condition were able to conduct the task wherever it suited them; feedback from participants after the experiment was that students did the task in their residence by themselves. As such, both the supervised and unsupervised tasks were conducted independently and individually.

The instruction packages (Appendix 2) and card sort tasks were identical for all four groups and the same amount of information was provided no matter the condition. The random assignment minimized the effects of response bias by providing equal chance of assignment to different group for early or late registrants. Sampling was done until at least 10 students were assigned to each condition for a total of 42 participants. For consistency, the same researcher, who was known to all participants before the experiment, supervised all manual and online observed card sorts in the same physical location. In addition, to gain familiarity with the task, there was a short three-construct, ten-item closed card sort used to introduce the card sort process.

In order to avoid the student focusing on the card sort task, deception was used in describing the purpose of the study. It was initially explained to the students that the survey would be used to identify key factors that influence an adolescent’s decision to attend university and the expectations of students following graduation. It was explained that, as part of the research, student participants were being involved in instrument development by conducting a card-sorting exercise. Following data collection, the true intent of the experiment was revealed to the students and they were given the option to withdraw from the study, though none did.

Experiments can be assessed on four major types of validity – statistical conclusion, internal, construct, and external (Shadish et al, 2003). Each of these validities was addressed through experimental design. Statistical conclusion validity refers to inferences regarding correlation between treatment and outcome. Validity is assessed through determination of statistical power of tests for the selected sample size (Shadish et al, 2003). The sample of at least ten students per group indicates the appropriateness of non-parametric procedures. Non-parametric testing of the difference between conditions was done using the Mann-Whitney U test, which can be used for samples sizes as small as five and does not assume normality of the underlying distribution (Kazmier, 1995). In the aggregate, there were sufficient observations (42) to conduct parametric analyses such as ANOVA and regression, though we recognize that the statistical power of these tests may be limited. Wood and Wood (2008), in reference to information architecture, held that 25-30 participants in a card sort will yield similar results to samples in the hundreds as long as the participants are representative of the actual users. Given that students are the actual users of university education, this sample size was deemed appropriate for a card sort.
Internal validity refers to whether the observed covariation between the presumed treatment and the presumed outcome reflects a causal relationship (Shadish et al., 2003). The major threats to internal validity are reduced through the use of random assignment to groups. As students registered for the study, they were assigned a random condition and provided that specific card sort to complete. While this resulted in unequal group sizes, it provided the greatest internal validity and was a consciously chosen approach.

Construct validity refers to whether the operationalization of the construct reflects the intended definition of the constructs (Shadish et al., 2003). The manipulation check is designed to ensure that the treatment is the intended manipulation. Students were asked to provide feedback on the card sort and in their responses, all focused on the survey topic of why they chose a university education as opposed to the method of card sort in their commentary.

External validity refers to the whether the relationship observed is generalizable over variations in persons, setting, treatment variables and measurement variables (Shadish et al., 2003). External validity is a perennial weakness and criticism of laboratory experimentation (Benbasat, 1990). While student-participant experiments are particularly singled out as being non-generalizable, the suitability of using students as representative participants for analysis tasks has been argued for (Abdel-Hamid et al, 1999) and against (Marakas & Elam, 1998); however the aim of the experiment argues for their use. Students are frequently used for card sorting activities (Bassellier & Benbasat (2007); Campeau et al. (2007); and Brown et al. (2016)) and the experiment was designed to occur in the domain of their experience – gaining a university education. The trade-off between external validity and the preceding other three types of validity was, therefore, consciously made.

Finally, there were different card sorting software options available for use. In a study that looked at the importance of ease of use for participants and researchers, Chaparro (2008) noted that researchers preferred WebSort to set up and analyze a card sort and thus WebSort was used. Websort duplicate the manual card sort process by providing a list of items which can be dragged and dropped onto “piles” of categories.

**Results**

A total of 42 card sorts were conducted, with 12 in Online Unobserved and 10 in the remaining three combinations. Each card sort was evaluated using Cohen’s Kappa, a measure of inter-rater agreement comparing the researcher-defined item-category matches to the participant. Key statistics are in Table 1. Because certain non-parametric methods are used, both mean and median, score and rank are reported.

<table>
<thead>
<tr>
<th></th>
<th>Mean Score</th>
<th>Rank</th>
<th>Median Score</th>
<th>Rank</th>
<th>High Score</th>
<th>Rank</th>
<th>Low Score</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Observed</td>
<td>0.547</td>
<td>16.4</td>
<td>0.628</td>
<td>13.5</td>
<td>0.734</td>
<td>2</td>
<td>0.227</td>
<td>39</td>
</tr>
<tr>
<td>Manual Unobserved</td>
<td>0.533</td>
<td>18.7</td>
<td>0.564</td>
<td>19.0</td>
<td>0.694</td>
<td>4</td>
<td>0.216</td>
<td>41</td>
</tr>
<tr>
<td>Online Observed</td>
<td>0.537</td>
<td>21.0</td>
<td>0.595</td>
<td>20.0</td>
<td>0.774</td>
<td>1</td>
<td>0.329</td>
<td>35</td>
</tr>
<tr>
<td>Online Unobserved</td>
<td>0.415</td>
<td>28.5</td>
<td>0.404</td>
<td>30.5</td>
<td>0.642</td>
<td>9</td>
<td>0.201</td>
<td>42</td>
</tr>
</tbody>
</table>

As the null-hypothesis would suggest no difference in scores across categories or experimental conditions, we first conducted an ANOVA to determine if there were differences between the groups in the sample. At p=0.09, we can reject the null hypothesis and find that there is a difference between the manual versus online and observed versus unobserved card-sorts.

The next analysis was conducted to the hypotheses as summarized in Table 2. Mann-Whitney U tests were conducted for each permutation. Tests for expectations of no difference in accuracy were two-tailed; tests for theorized expectations of difference were one-tailed (* p≤0.05, † p≤0.10).

The final step was to identify a potential model for the differences between the approaches to discern the relationship. We conducted two different regression analyses, one entering terms and one step-wise, coming to two complementary results.

We first report the correlations between conditions in Table 3.
Manual versus Online Card Sorts

Hypothesis  | Significance  | Support  
--- | --- | --- 
H1 Mn=OL | 0.047 * | Not Supported 
H2 Ob>Un | 0.081 † | Weak Support 
H3a OO=MO | 0.384 ns | Supported 
H3b OU=MU | 0.070 † | Not Supported 
H3c MO>MU | 0.337 ns | Not Supported 
H3d OO>OU | 0.069 † | Weak Support 
H3e MO>OU | 0.016 * | Supported 
H3f OO>MU | 0.337 ns | Not Supported 

MN= Manual  
OL = Online  
Ob = Observed  
Un = Unobserved  
MO = Manual observed  
MU = Manual unobserved  
OO = Online observed  
OU = Online unobserved

Table 2. Mann-Whitney U Hypothesis Tests

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Manual Observed</td>
<td>0.202 †</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Manual Unobserved</td>
<td>0.132</td>
<td>-0.313 *</td>
<td></td>
</tr>
<tr>
<td>3. Online Observed</td>
<td>0.076</td>
<td>-0.313 *</td>
<td>-0.313 *</td>
</tr>
<tr>
<td>4. Online Unobserved</td>
<td>-0.386 **</td>
<td>-0.354 *</td>
<td>-0.313 *</td>
</tr>
</tbody>
</table>

Significant at † ≤0.10, * ≤0.05 and ** ≤0.01

Table 3. Observed Correlations

The strongest correlation was a negative one between the card sorting accuracy score (Cohen’s Kappa) and the use of the Online Unobserved approach. Conversely, a weaker but still observable positive correlation with card sorting accuracy was noted with the Manual Observed approach.

In examining the two regression models (Table 4), a common theme can be noted. In Model A, the online-unobserved approach is used as the baseline and each of the other three approaches provided a significant performance improvement above the baseline. In Model B, the other three approaches are used as the baseline and the online-unobserved approach provides a significant performance reduction from it.

<table>
<thead>
<tr>
<th></th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.415 (0.046)</td>
<td>0.555 (0.028)</td>
</tr>
<tr>
<td>1. Manual Observed</td>
<td>0.159 * (0.068)</td>
<td></td>
</tr>
<tr>
<td>2. Manual Unobserved</td>
<td>0.139 * (0.068)</td>
<td></td>
</tr>
<tr>
<td>3. Online Observed</td>
<td>0.122 † (0.068)</td>
<td>-0.140 * (0.053)</td>
</tr>
<tr>
<td>4. Online Unobserved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.155</td>
<td>0.149</td>
</tr>
<tr>
<td>Adjusted R-Squared</td>
<td>0.089</td>
<td>0.128</td>
</tr>
<tr>
<td>No. Observations</td>
<td>42</td>
<td>42</td>
</tr>
</tbody>
</table>

Table 4. Regression Models

Discussion and Conclusions

Where Harper et al. (2003, p.2049) found that “a computerized card sorting task is just as effective at eliciting knowledge as a manual card sort”, our findings suggest that there is a significant difference between manual and online card sorts. Because our hypothesis that manual and online card sorts were equivalent (H1) was not supported, it would be problematic for researchers to assume equality of methods. Similarly, our hypothesis that observed card sorts were more accurate than unobserved card sorts (H2) was weakly supported affirms the findings of Adair (1984), McCambridge et al. (2014), and Sedgwick and Greenwood (2015) that observation can impact performance. If observation can impact the results, researchers should not mix observed and unobserved approaches without realizing the potential impact on their study.

It should be noted that for the defacto norm of observed card sorts (i.e. conducted in a lab on-site, either manually or on a computer), there is no significant difference between manual and online observed (H3a). If the two methods were compared under these conditions, then the literature regarding the similarity in
results for online and manual card sorts is supported. This is not the case if one were to assume that manual and online unobserved (i.e. done in a place of work or residence) would be identical because there is an observed difference between the two approaches (H3b) and supports the findings of Greve (2014).

If one were to visualize card sort accuracy on a scale, manual observed would be followed by almost identical manual unobserved and online observed and then, at a distance, by online unobserved. This is supported by the lack of significant difference between manual observed and manual unobserved (H3c) and manual unobserved and online observed (H3d), the weak significance between online observed and online unobserved (H3e) and the strong significance between manual observed and online unobserved (H3e). The last of these observations is the most problematic because, in terms of scale development procedures, a researcher may conduct a manual card sort with those who are close at hand but send a link to an online card sort to those further afield based on the assumption that they are equally accurate.

Because of the importance attached to methodology when conducting a study, IS researchers must be mindful of their design of the card sort and be consistent in their application of it. Essentially, should researchers find themselves with both proximate and distant candidates for the card sort, they should ensure the same card sort technique is used. The key issue is the difference between manual observed and online unobserved because these would be the default settings under these conditions. Specifically, if researchers have participants in the local area, they would likely use a manual card sort; however, if they have someone further away, they would likely use an online card sort. This situation should be avoided. Rather, if researchers have any participant in the proposed sample at a distance and require an online unobserved card sort, then all participants should be given an online unobserved card sort.

As with all studies, this one has limitations. The first limitation of this study is the sample size, leading to non-parametric measures. However, as a cautionary tale where the intent is to warn researchers of potential issues rather than fully quantifying the degree of risk, we believe that the sample size is appropriate for the analysis and task. A second limitation is in using a single card sort with only one level of complexity. However, we developed a complex card sort task with the intent of drawing out the differences in the conditions because we expected that a participant might overcome the noted effects in a simple task. A third limitation is that the study only looks at accuracy of the card sort. Besides assessing reliability of the instrument, there are other reasons for conducting card sorts including soliciting refinements to items, etc. We were only focused on examining issues surrounding the conditions under which card sorts occur, not their ultimate value, which we acknowledge and support.

In conclusion, this cautionary tale demonstrates that the use of cards sorts for IS research, coupled with the proliferation of online card sorting software, means that researchers need to be aware that their studies can be affected when manual observed and online observed card sorting techniques are used concurrently. Mindfulness of the impact of card sort design can improve the consistency of their results.

**REFERENCES**


Appendix 1 – Extract of model construct and card sort items

1. Personal Factors. Characteristics in the participant’s life or environment that could influence the decision making process.

Item Academic Characteristics
PF-A1 Province of entry
PF-A2 Final year high school/CEGEP average
PF-A3 Years of college/university before attending UNIVERSITY X

Item Family Characteristics
PF-F1 Father’s highest education
PF-F2 Mother’s highest education
PF-F3 Family income level

Item Individual Characteristics
PF-I1 Gender
PF-I2 Age
PF-I3 First official language

2. Decision Factors. Individual and personal considerations when deciding to go to university in general and UNIVERSITY X in particular.

Item Academic Potential
DF-A1 My marks affected where I thought I could apply to school.
DF-A2 I had many different options for going to university based on my marks.

Item Personal Preference
DF-P1 UNIVERSITY X was my first choice university.
DF-P2 I was accepted at many universities.
DF-P3 I applied to many universities.

Item Financial Considerations
DF-F1 I was offered scholarships at many universities.
DF-F2 I had the financial means to go to any university.
DF-F3 I chose UNIVERSITY X as I could get paid while I went to school.
DF-F4 I selected my school based on cost.

3. University Decision. Reasons for deciding to go to university in general.

Item Educational Considerations
DU-E1 Get a good general education.
DU-E2 Prepare for graduate or professional school.
DU-E3 Increase knowledge in an academic field.
DU-E4 Become a more educated person.
DU-E5 Develop critical thinking and reasoning abilities.

Item Job-Related Considerations
DU-J1 Prepare for a specific job or career.
DU-J2 Get a good job.
DU-J3 Obtain a university degree for entry into the workforce.
DU-J4 Discover career interests.
DU-J5 Develop a broad base of flexible skills.
Appendix 2 – Card Sort Instructions

1. In this card sort, you will be given a number of descriptive questions and the key categories of the research model. Your objective is to put the descriptive questions into the categories provided using the online software.

2. On each card a single statement is written. Sometimes, multiple cards contain statements about similar things or statements that reflect an underlying concept. In addition to the cards, there are envelopes with labels that represent the categories of the model. Two additional envelopes titled “does not fit any category” and “does not make sense/are confusing” are also provided.

3. Your task is to sort the index cards based on the similarities and differences among the statements on each card. Each envelope reflects one underlying concept. The titles are provided, you need only put the descriptive cards into the envelope provided.

4. The following are additional guidelines for the procedure:
   • You should place the cards into groupings such that each group of cards relates more to one another than they do to cards in a separate category. Many statements will appear to be very similar, so try to determine the underlying idea each statement reflects.
   • Separate groupings should not be used based on different degrees of the same concept. Instead, different degrees of the same concept should be grouped together. For example, the words “hot” and “cold” could be grouped together if the underlying concept is “degree of heat”. “Wet” and “dry” could also be grouped together if the underlying concept is “level of moisture”.
   • Separate groupings should not be based on how you would answer the question. For example, all statements that you agree with or believe to be true should not be placed together. All statements that you disagree with or believe to be false should likewise not be placed together.
   • The number of cards in each category is not necessarily the same and could vary extensively depending on the category.
   • You will also see two additional envelopes for any statements that “do not fit any category” or “do not make sense/are confusing”. Place cards into these categories as you see necessary.

5. Complete the sorting task by:
   • reading all of the statements,
   • grouping the statements in the categories according to underlying concepts,
   • re-sorting the cards, re-organizing the piles as necessary if you change your mind,
   • reviewing each grouping one more time once you have established your final sorting to ensure that all the cards are in the “right” category and that none of the statements fit better in a different category, and
   • put the cards into the appropriate envelope.

6. This is not a test. The purpose of this sorting exercise is to see whether or not the statements relate to each other as originally intended or if further work is needed.