

Making sense of ‘pure’ phenomenography in information and communication technology in education

Graham Alsop and Chris Tompsett*

Kingston University, UK

Research in information and communication technology in education places an increasing emphasis on the use of qualitative analysis (QA). A considerable number of approaches to QA can be adopted, but it is not always clear that researchers recognize either the differences between these approaches or the principles that underlie them. Phenomenography is often identified by researchers as the approach they have used, but little evidence is presented to allow anyone else to assess the objectivity of the results produced. This paper attempts to redress the balance. A small-scale evaluation was designed and conducted according to ‘pure’ phenomenographic principles and guidelines. This study was then critiqued within the wider context of QA in general. The conclusion is that pure phenomenography has some procedural weaknesses, as well as some methodological limitations regarding the scope of the outcomes. The procedural weaknesses can be resolved by taking account of good practice in QA. The methodological issues are more serious and reduce the value of this approach for research in collaborative learning environments.

Introduction

This paper presents a critical review of phenomenography as a qualitative research process for use within information and communication technology in education (ICTE). In most published studies that have used this approach to qualitative analysis (QA) the emphasis is on the conclusions reached, and little consideration is given to the research process or to factors that might restrict the validity or generalizability of the conclusions.

This paper seeks to redress the balance by presenting a ‘pure’ phenomenographic study and a critique of the research process that was applied. This critique raises some

*Corresponding author. Learning Technology Research Group, Kingston University, School of Computing and Information Systems and Mathematics, Penrhyn Road, Kingston upon Thames, Surrey KT1 2EE, UK. Email: c.p.tompsett@kingston.ac.uk

significant problems that need to be addressed if pure phenomenography is to be used for future research in ICTE.

The most detailed level of analysis (the evaluation study) is focused on the use of software that was developed within the NetPro II project¹ to facilitate group coordination and management for problem-based learning. The purpose of the evaluation was to provide qualitative feedback to inform the next stage of design and development of the software. The design team was particularly interested in how the students perceived the tool, as opposed to whether the students understood or shared the designers' perspective. From a range of approaches that were available, phenomenography appeared well suited given the limited period of time during which data could be collected. The possibility of producing 'results' more quickly than with other approaches to QA was a further advantage. This study followed, as far as possible, the pure phenomenographic process outlined in Marton and Booth (1997).

The critique of this research (the process study) was based on a quasi-teachback process (Kidd, 1987) between the researcher on the evaluation study and a second researcher who had not taken part in the research up to this stage. In this case, the second researcher acted as an expert in qualitative analysis, and the teachback activity focused on elaborating the evaluation study within this wider research framework. The format of this critique allowed the research process to be explored in detail, and emphasized, in particular, those aspects that are distinctive to phenomenography rather than common across QA.

This critique revealed a number of weaknesses that need to be addressed in any future use of phenomenography in ICTE. At the simplest level, the process needs to be revised to ensure that the method can track the extent to which the researchers' own values influence the interpretation and analysis of the data. From a methodological point of view, it must be recognized that the principles that underlie phenomenography can only produce a narrow, snapshot model of what understanding might be, and provide little insight into learning.

The first two sections of this paper provide an introduction to phenomenography and the evaluation study. The central part of the paper covers the process study and concentrates on four critical differences between pure phenomenography and other approaches to QA. Three of these are reviewed within the context of the evaluation study and challenge the validity of this approach as it is currently presented. The fourth arises from the process study itself and suggests that there could be strong limitations to the scope of any phenomenographic research. The conclusions to the paper integrate these four criticisms and suggest that pure phenomenography, at least in ICTE, can only improve by becoming less phenomenographic.

Phenomenography

Phenomenography is presented as a model of qualitative research and analysis that is distinct in its principles, its focus, its methods and its outcomes. The term phenomenography was first defined as a distinctive approach to qualitative research and analysis by Ference Marton in an article in *Instructional Science*, in which he comments:

The kind of research we wish to argue for is complementary to other kinds of research. It is research which aims at description, analysis, and understanding of experiences; that is, research which is directed towards experiential description. Such an approach points to a relatively distinct field of inquiry which we would like to label phenomenography. (Marton, 1981, p. 185)

This approach evolved over the previous five years in a number of studies in education that each focused on learning and understanding within a specific subject (e.g. economics; Dahlgren & Marton, 1978), and even on a specific task within that subject (e.g. understanding a specific text; Marton & Säljö, 1976a, b). The complementary nature of this approach, for Marton, is established by distinguishing phenomenography from other forms of enquiry.

Phenomenographic studies are strictly empirical and non-constructivist (Svensson, 1997, p. 164). Despite the alignment with the empirical tradition, phenomenography must be distinguished from both conventional science and educational psychology. It is distinguished from science, as there is no intention to describe an objective world that is independent of individuals (Marton & Säljö, 1976a). It is distinguished from educational psychology, as there is no intention to produce a model of the capability of humans to learn, perceive and/or behave without reference to a specific context (Marton & Booth, 1997). In contrast, phenomenography studies the inter-relationship between the individual and the objective world, and draws conclusions about how individuals 'conceive of various aspects of their reality' (Marton, 1986, p. 42). As such, phenomenography is concerned to understand the limitations to the way in which a specific aspect of the world can be experienced by individuals (Säljö, 1996, p. 12). Thus, Dahlgren and Marton's study of students' understanding in economics should inform us about the different ways in which that part of economics could be perceived by any similar student, but would be irrelevant to understanding how the same students understand a different subject.

Phenomenography must also be clearly distinguished from phenomenology, which Marton considers to be unnecessarily abstract (Marton, 1981). Phenomenography does not accept that it is possible to separate 'that which is experienced from the experience per se' (Marton, 1981, p. 180), while phenomenology is concerned to understand how a subjective perception of 'essence' can be understood as distinct from particular experiences. In addition, where phenomenology is limited to the 'pre-reflective level of consciousness ... of the taken-for-granted world', a phenomenographic study includes 'both the conceptual and the experiential', both 'what is thought of' and 'that which is lived' (Marton, 1981, p. 180). Where phenomenologists disagree (Hasselgren & Beach, 1997), they argue against, rather than for, pure phenomenography.

In later papers, the prefix 'pure' is added to the term phenomenography to denote the research process that was originally defined (Marton, 1986). This emphasizes the generic nature of the research process and distinguishes it from other studies that share the same educational focus as the original studies, irrespective of the approach that is taken (Svensson, 1997, p. 164). For the remainder of this paper, phenomenography, unless indicated otherwise, is used to refer to the process of research and

should be read as 'pure' phenomenography. We shall also need to restrict our attention to one definition, and we concentrate on Marton's view within the remainder of the paper. Although some other approaches have been proposed (Hasselgren & Beach, 1997) these do not appear to have found favour within research in ICTE.

These distinctions provide a set of principles that establishes phenomenography as a qualitative approach to research that produces statements about how individuals can describe, or equivalently be aware of, particular experiences, but phenomenography is not unique in this. Other approaches to QA, based on different principles (e.g. action research, discourse analysis, ethnomethodology, grounded theory, etc.), would share the same interest in describing the ways in which individuals can understand their experience of a phenomenon in a common 'outer world'. Phenomenography distinguishes itself from each of these in practice through at least one of three characteristics that we consider in the remainder of this section:

- the presumed objectivity of data collection;
- the structure of the outcome space as a hierarchy; and
- the characterization of this hierarchy as a limit to the experience of any individual.

In conventional phenomenography, qualitative accounts are collected, in a one-pass approach, from a sample of subjects who share experiences of a particular phenomenon. The typical approach for adults is to adopt a semi-structured interview with a well-structured sample, but alternative approaches are possible (Marton & Booth, 1997, pp. 131–132). The suggestion (Booth, 1997, p. 130) that a 'theoretical sample' should be used, as in grounded theory, is not possible until the later stages of a cyclical process of data collection. Sampling should be designed to capture diversity rather than to produce a statistically balanced representation. Semi-structured interviews add depth to the data, but this requires sensitivity to avoid the interview becoming a 'diagnostic discourse' (Marton & Booth, 1997, p. 130) rather than an exposition of the subject's own perception. Analysis of these accounts is then treated as a distinct stage. Few other approaches to QA would consider these accounts to be so unaffected by the collection process.

The outcome space is characterized as a hierarchically structured, multi-dimensional super-set of descriptions, where each subcomponent is a multi-faceted issue or aspect bounded by a finite range of values. This needs to be considered in two ways.

Firstly, we can consider the outcome space as the endpoint of an empirical process. Since almost all other approaches to QA produce a similar super-index for the language and terminology used by subjects, the major difference is that phenomenography stops at this point. As an empirical analysis should be traceable, each part of the hierarchy acts as a cross-reference to the original accounts from which it stems (Säljö, 1996). Entwistle (1997) clearly recognizes that further processes are necessary and outlines some principles for how a phenomenographic study should progress beyond this. Other models of QA would certainly do so. Grounded theory (GT) develops an axial model to represent the sequential aspects of an experience or decision (Glaser & Strauss, 1967), while other approaches would be more concerned

to explore normative aspects through consideration of frequency of recurrence, and so forth. Despite Entwistle's suggestion, phenomenographic studies are finished with the production of the outcome space.

Secondly, the outcome-space is not just a super-index to the original accounts. As Marton notes, 'The set of categories is thus stable and generalizable between situations, even if the individuals "move" from one category to another on different occasions' (1981, p. 194). In other words, for any well-defined phenomenon, there is a fixed number of ways of conceiving of that reality (Marton, 1986, p. 42). This provides strong constraints on how reality could be perceived, but within these places no limitations on how individuals perceive any phenomenon on any occasion. A more detailed account reflects a deeper awareness of that particular phenomenon at that time, but that is a statement about the account, and not about the individual themselves. More complex experiences will reflect a higher level, or a more 'authorized' view of the world. In education this will normally be what the students should have learned (as with deep/shallow learning). Outside education, the 'authorized' view might reflect a higher level of scientific development, or more advanced cultural development (Marton, 1981, p. 184).

Despite the significance of this super-index to phenomenography, relatively little is written about the process by which it is derived—a process that is acknowledged to be a '... discovery procedure which can be justified in terms of results, but not in terms of method' (Marton & Säljö, 1984). Instead, the phenomenographic researcher is expected to gain an understanding of the process from reading a sufficient number of case studies. Even so, apart from a description in that work, the case studies, as published, provide little detail, whether the studies are from education (for example, Booth, 1997; Svensson, 1997; Entwistle, 1997; Marton & Pang, 2003) or from more diverse fields—for example, teaching and cultural analysis (Mugler & Landbeck, 1997), public policy (Irvine, 2002), and nursing (Widang & Fridlund, 2003).

In summary, phenomenography presents itself as a powerful research tool that produces an objective, qualitative description to represent the way that individuals perceive reality. For researchers in ICTE, it offers the possibility of encapsulating the different perceptions of particular systems that are held by the designers, the educators and the learners, with relatively limited cost in research time. However, in doing so, some critical differences exist between phenomenography and other approaches to QA. The evaluation study and the process study that follows allow these differences to be explored in practice and in theory.

The next section covers the use of phenomenography in the most interesting strand of the evaluation study. By choosing to do so, we can exclude aspects of the case study that are too contextualized, and can retain sufficient details to illustrate the process and support the critique of the process that follows within the remainder of the paper.

The evaluation study

As already noted, the evaluation study was designed to report on the experience of teams of undergraduate students who were expected to collaborate on a multi-media

design task. In this case, the students were on a multimedia design module at Level Two in higher education.

Two software systems were available to support the students: the system that was the focus of the evaluation (the tool), and the university's learning-management system. The tool provided support for coordinating work, collaboration in the production of interim designs, and submitting work to deadlines. For academic staff, the tool allowed the progress of each team to be monitored efficiently and feedback to be given on the various components of the assignment as they were produced. From the designers' perspective, there was a particular interest in the extent to which the students conceived of the tool as instrumental in supporting collaborative working. The steps of the evaluation study will now be described in sequence.

Evaluation study: data collection

A number of factors restricted the options for data collection. Students needed to have had sufficient opportunity to use the tool and deliver designs to the deadlines, but their time was severely constrained, at that point in the semester, by the demands of other studies. Students needed to be sufficiently willing to provide feedback, but, ethically, they could not be required to do so. For both these reasons, self-selection was agreed as the only option (and was taken up by 24 students from a total of 111). The period over which data were collected and the time required from each student was limited as far as possible. The data were collected as written responses to a single prompt (Alsop & Tompsett, 2002) by the module leader.

The prompt did not make direct reference to collaboration:

We would like you to concentrate on what you, personally, could identify as a single occasion which you consider as the occasion (or one of the occasions) which was the best educational experience when using NetPro or Blackboard.

We would like you to tell us about this event. Please write about 8–15 sentences in the space below that outline the details of how the event occurred.

In comparison with semi-structured interviews, the accounts are collected in parallel and without any further prompts. Each account is written and includes as much information and detail, or as little, as the subject chooses. Both factors should increase the objective nature of the data as individual accounts. Students do not have any opportunity to influence each other, and the opportunity for the researchers to influence the views of the subjects is limited. The subjects are familiar with presenting a position as a written account and so, despite the lack of any interactive 'discussion', this approach would still be classified as a 'discursive' model of data collection in phenomenography (Hasselgren & Beach, 1997, p. 196). The collection of accounts in parallel eliminates the risk that details, which have been presented in early interviews, might influence the pattern or detail of subsequent interviews.

Although the data were not collected conventionally, the accounts that were collected showed sufficient variation in length, detail and content to support the phenomenographic analysis that followed. In particular, from the designers'

perspective, although no direct reference had been made to collaborative learning within the prompt, collaboration did form an integral aspect of the students' accounts of using the tool.

Evaluation study: analysis

Each account is one description of one experience, which is limited by what was perceived by the individual at the time and considered to be relevant on this one occasion. Phenomenographic analysis starts with these accounts as a 'pool of experiences' and develops a single 'stripped' description: 'in which the structure and essential meaning of the differing ways of experiencing the phenomenon are retained ...' (Marton & Booth, 1997, p. 114). Two principles guide the analysis. Firstly, that the details within the accounts will be hierarchically structured (Booth, 1997, p. 138). Secondly, it is argued that the model should be as 'parsimonious' as possible (Marton & Booth, 1997, p. 125).

The model that was designed for the evaluation study is principally based on Marton and Booth (1997, p. 114), with Marton and Säljö's (1984) work providing additional clarification. The process worked through three phases of analysis, each of which could have triggered a reassessment of an earlier phase:

1. structured reading: reading and re-reading all the experiences a number of times to identify the key aspects/issues of a phenomenon;
2. identifying variation for each aspect/issue: reading the relevant cases to identify the possible variation in the way this is experienced; and
3. structuring experiences:
 - (a) separating into levels if possible, and
 - (b) clustering into an outcome space that is hierarchically structured.

(Note: The texts use the term aspect at both a macro and detailed level of analysis. In this paper the word issue is used at the highest level, and aspect for lower levels.)

Phase 1: structured reading. The relationship between the researcher and the experiences collected from the subjects is considered to be that of an independent observer. The researcher is expected to 'step back consciously from her [sic] own experience of the phenomena and use it only to illuminate ways in which others are talking of it, handling it, experiencing it, and understanding it' (Booth, 1997, p. 121). A separation between data collection and analysis reinforces this, and emphasizes a model of research in which the accounts are treated as objective data. In the structured reading phase the complete set of accounts is read and re-read a number of times before any attempt is made to make notes or begin a more formal analysis, even if this does not always appear to be followed (for example, Ramsden *et al.*, 1993). It is argued that any analysis before re-reading may 'fix' the analysis on aspects, or details, that appear important to the analyst in the first few cases, but that would not be significant after reading the full set. In the evaluation study the full set was read three times.

At the end of this phase, two issues became evident: issue A, submitting work; and issue B, differences between the two technical systems. The next phase of analysis cycles through each issue in isolation from the others.

Phase 2: identifying variation. This phase represents a shift from identifying holistic issues to capturing significant variation for one issue. On each cycle (a ‘freeze’; Marton & Booth, 1997, p. 133), comments that are not relevant to that issue are ‘ignored’. The criterion of parsimony requires that ‘similarity of view’ is captured by the researcher within a single representative phrase or statement (termed comments from this point onwards). Repetition of the same view is then ignored; phenomenography is intended to define the limits of how a phenomenon is experienced rather than what is normative. Each of the comments that remain can be traced back to at least one account, but is now considered against the different comments that could be made about the same issue. If the sample is selected appropriately this will represent all the possible comments that can be made (Marton & Booth, 1997, p. 120). If the accounts are processed in their original sequence and the comments retained in the sequence in which they are first noted, this restricts the opportunity to over-interpret the data at this point.

For issue A, submitting work, there was little difficulty in identifying relevant comments by elimination; comments were excluded if they were restricted to the learning-management system or made a direct comparison of common features. Table 1 provides the first list of comments (as noted and written by the original

Table 1. Issue 1: submitting work—separate comments, in order of first occurrence

Issue A, submitting work	
a	Loading a document up
b	Loading a document up and getting a smiley face ^a
c	Loading an empty document up and getting a smiley face
d	Concerns that they could see each others’ work
e	Concerns about them not sharing the work among each other
f	Concerns that not all subjects look at the upload of information from a sharing point of view
g	Positive identification that the exchange of ideas was available and possible
h	Suggestion that security should not allow sharing by default—use passwords
i	An advantage is that others can see your work
j	People can copy
k	Where are the grades?
l	Doesn’t track changes in documents
m	Could change the file type to PDF
n	People can use your work to get ideas
o	No feedback from staff
p	You can see what had not been done yet
q	More convenient to upload documents off site

^aA visual clue that a file had loaded successfully.

researcher) for this issue from the evaluation study, after the first stage of interpretive coding.

Although this list appears to cover a wide range of aspects that might appear disconnected in the table, each can be traced back to a link within the account of at least one student. The richness of this variation provides a *post-hoc* validation of the approach to data collection that was used. A more detailed level of coding is introduced in the following phase.

Phase 3: structuring experiences. The outcome of this phase is 'categories of description' and is well defined. The process by which this is achieved, as noted before, is unclear. Marton and Booth comment that these descriptions should:

as a rule, form a hierarchy ... defined in terms of increasing complexity, in which the different ways of experiencing the phenomenon in question can be defined as subsets of the component parts and relationships within more inclusive or complex ways of seeing the phenomenon. (Marton & Booth, 1997, p. 125)

They continue: 'The different ways of experiencing the phenomenon can even be seen as different layers of individual experiences'. From this, complexity appears to be as a property of an account that emphasizes the detail that is included, although this phase is based on the comments themselves. This suggests that certain comments would only occur within complex accounts—but this does not allow for inclusiveness. This seems to suggest that some comments would only occur if lower levels of experience had to be there, even if relevant comments have not been included in any one account.

In reality, the first part of phase 3, 'separating into layers if possible', involved two interlinked processes. At a macro level, comments were subdivided into four aspects: 'loading a document', 'managing files', 'sharing' and 'feedback'. These four aspects could be layered by exploiting inclusiveness. For example, none of the comments on sharing would make sense unless the subject was already aware that files could be managed within their own workspace. The reverse would not necessarily be true.

On a more detailed level, a local structure was based on a more systematic coding of closely related expressions and, where necessary, a reapplication of the principle of inclusivity. As a simple illustration of systematic coding we consider the four comments (a, b, c, q) from Table 1 that refer to 'loading a document'. Coding and sequencing across a small set is relatively unproblematic. Table 2 presents these, together with their coding in the sequence that was provided by the researcher. The numbers in the first column are taken from the final ranking that was produced. Specific issues in coding are now discussed.

With larger and more diverse sets, such as the eight comments that were identified as relevant to sharing, a more detailed analysis is necessary. Table 3 presents these comments arranged into six levels (6–11), again with the researcher's notes and coding. Figure 1 shows the aspect within a hierarchical structure. Although there is insufficient space to justify this sequence, it provides a good example to illustrate how complexity and inclusivity interact within a more structured coding system.

Table 2. Loading a document: sequenced in decreasing levels with structured coding

	Submitting work: loading a document	Final coding
4	q: More convenient to upload documents off site	[Loading: off-site]
3	c: Loading a document up and getting a smiley face	[Loading: confirmation]
2	b: Loading an empty document up and getting a smiley face	[Loading: confirmation and empty]
1	a: Loading a document up	[Loading]

In this case, the comments that discuss the exchange of files (levels 6 and 7) are set below those that refer, or imply, an exchange of ideas (levels 8–11). Even though none of the individual comments makes an explicit comparison of the two terms, the researcher considered that the exchange of ideas between teams in the workspace requires that the relevant files can be exchanged, but, as above, the reverse would not necessarily hold true. Within the set that covered the exchange of ideas, those that discussed sharing as ‘learning’ rather than exchange seemed to represent a deeper level of understanding of the purpose of the system. The comment at level 11, which implies that the ‘design’ of the tool was not fixed, also provides a good example where complexity sets this comment above the previous level.

Table 3. Sharing sequenced in decreasing levels with structured coding

	Initial coding	Researcher’s note (verbatim)	Coding
11	h: Suggestion that security should not allow sharing by default—use passwords	A wish to share work, but to have choice. Suggestion that security should not allow sharing by default—use passwords who to share it with	[Sharing: +ve, with choice, ideas]
10	g: Positive identification that the exchange of ideas was available and possible i: an advantage is that others can see your ...	Positive wishes to share work and learn from it	[Sharing: +ve, learning, ideas]
9	f: Concerns that not all subjects look at the upload of information from a sharing point of view	Recognition that not all view sharing with the same goals in mind	[Sharing: asymmetric, ideas]
8	n: People can use your work to get ideas	Recognition that sharing will aid the work of others	[Sharing: +ve, ideas]
7	m: Could change the file type to PDF	Opportunity to share work identified with the addition of some security	[Sharing: –ve, restricted]
6	d: Concerns that they could see each other’s work j: people can copy	Sharing work leading to concerns about plagiarism	[Sharing: –ve]

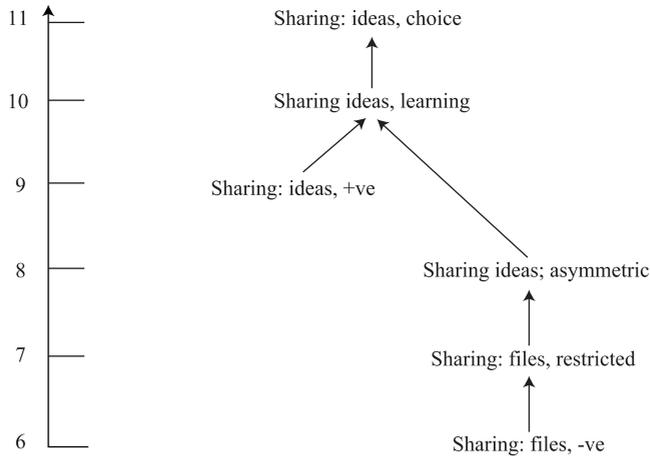


Figure 1. Hierarchical structure to sharing

In the hierarchical representation, higher levels either include new aspects (e.g. choice, learning) or replace one value with another. When particular values are considered to be juxtaposed, they appear on different strands of the hierarchy.

Once this process is completed for each aspect of the issue, then a complete hierarchy for this issue can be produced. An outline of the hierarchy for submitting work is provided in Figure 2. As the hierarchy for sharing is already included in Figure 1, only the first three layers of this aspect are shown.

At the final level, when integration occurs across all the issues, then the analysis is complete. Phenomenographers argue that this index gains predictive power: if a suitable set of accounts has been collected, then any future account must also conform to the same structure.

For the evaluation study, predictive power would be of little value. The outcomes of this study would be relevant to any users of the same software, but the intention of the designers was to alter the software and to change, and hopefully improve, the experience of the students. In such a situation the designer would, ideally, be able to predict what will stay the same and what will change. However, a phenomenographic study cannot make such a separation!

This does not reduce the potential of phenomenography as a research tool. The study provides evidence that the students can perceive the tool as instrumental to their learning. However, this potential could be questioned if the process is not as inherently valid as is claimed. It is this issue that is considered in the process study.

Process study

A teachback exercise is based on Pask's conversation theory (Pask, 1977), and requires a 'novice' to coach an 'expert' through solving a problem. This allows the 'expert' to explore the layers of understanding of the novice, beyond the ability to produce the solution itself. In our case the original researcher coached a second

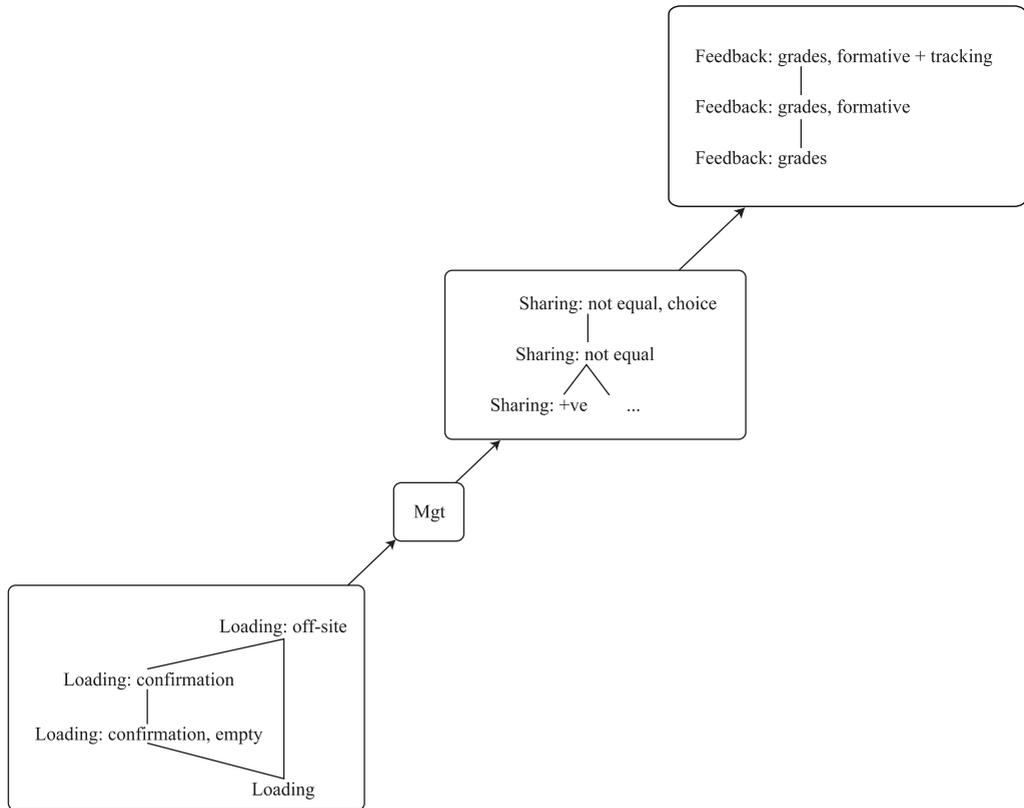


Figure 2. Outline of outcome space for issue A, submitting work

researcher, acting as an expert in QA, through the evaluation of the case study as a particular example of qualitative research. The second researcher provided a critical framework within which the key distinctions between phenomenography and other approaches to QA were highlighted. When a more detailed understanding of phenomenography was required, additional sources were checked (for example, Marton, 1994), but, in terms of pure phenomenography, these either confirmed what was written elsewhere or veered from the guidelines that would appear to distinguish pure phenomenography from other approaches to QA.

The teachback exercise eliminated discussion of what would be sound practice in any approach to QA, but this section can only act as a summary of the many cycles of exposition, explanation, challenge, reading and re-reading of the literature that were involved. We focus, in particular, on three key differences between phenomenography and other approaches to QA:

- The collection of data.
- The independence of the outcomes from the researcher.
- The generalizability of the outcomes.

As we consider each of these, we attempt to understand whether these differences should occur, 'on principle', from the underlying model of knowledge (as with the first) or whether it occurs, *de facto*, from a difference in practice (as with the last). We conclude this section with some concerns, which are raised from the process study.

Collection of data

Two inter-related factors are distinctive in data collection: single-pass collection, and the independence of data collection from analysis.

Single-pass data collection, collecting all the data before analysis begins, is atypical in QA where a grounded model is developed. The starkest contrast to the single-pass model would be GT. From a GT perspective, any interaction with subjects is interpreted through the mind of the researcher—so data collection entails analysis. In almost all other cases, data collection and analysis are recursive. In any of these approaches the researcher should be self-aware (Ashmore, 1989, p. 32) and document the progressive interaction of their personal understanding of the phenomenon and analysis as field notes. In phenomenography the opposite is the case: data are collected as objective accounts, with no suggestion that subsequent interviews or studies would be needed.

This view of data collection is a direct consequence of the model of awareness that underpins phenomenography. Each subject develops their personal understanding of a phenomenon through an unstructured sequence of experiences, but this awareness cannot be observed. All that is collected for analysis by the phenomenographer is a set of individual accounts of similar experiences. Each account is bounded by what the subject is aware of on one single occasion, and this account cannot be interpreted as a limit on either what the individual could be aware of in the future, or, indeed, might have been aware of in the past.

If data collection allowed an account to be revisited in order to 'add' to it, then this would imply that the subject's account needed to conform to some standard model. Separating data collection from analysis ensures that this cannot take place and provides 'empirical data' for the analysis that follows. Even if additional cases were to be collected, the protocol to be followed for any new case could not be changed. However, objectivity in data collection can only ensure that the first stage of research is independent of the researcher.

Independence of the outcomes from the researcher

For almost all other approaches to QA, it is presumed that the analytical phase must be dependent on the researcher even if it is then argued that the conclusions become independent. In such cases, it is expected, as intimated by Entwistle (1997), that additional conditions must be met, as with 'theoretical saturation' in GT (Glaser & Strauss, 1967).

The 'in principle' justification for this difference appears, on first consideration, to be promising. There is an evident similarity between the development of understanding

in a subject, and the development of an outcome space by the analyst. Each learner develops an understanding of a phenomenon through a number of different experiences, and a description of one experience is captured, in a structured format, and treated as objective data. Similarly, the researcher develops an understanding of 'reading an account' through experiencing 'reading an account' a number of times. This researcher's understanding is then captured, in a structured format, as the outcome space and considered to be objective. However, the similarity breaks down when considered in more detail. The data that are collected from each subject is an individual account. Different subjects would be expected to produce different accounts even if they chose to describe the same experience. On the same basis, the outcome space that is produced by one researcher is a representation of their understanding, and different researchers should expect to produce different outcome spaces (see R-reference; Ashmore, 1989, p. 32). Far from justifying independence (Marton & Booth, 1997, p. 125), the 'in principle' argument suggests that the outcome space should depend on the researcher.

If there is no additional test to justify independence, and the 'in principle' argument fails, we are able to consider the possibility that this is established through practice in the specific case of the evaluation study. To do so we would need to consider each phase in turn, although the first two are relatively unproblematic.

Phase 1, structured reading, is distinctive from other approaches to QA as there is a specific requirement to 'hold back' from interpretation, but this particular phase can be justified 'in principle'. The phenomenographic model of understanding does not require individuals to reflect on each experience, but captures the response to a specific example after understanding has developed. Structured reading reflects the same principle. Phase 2 is common to all forms of QA. Each will require that the language, or rather the data, is collected, and converted into a more restricted code (e.g. sharing, +ve, choice, ideas). At this point, any comment or term in the code acts as a denotational marker that points to some text within one or more accounts. When taken in isolation, any comment, such as 'Loading an empty document up and getting a smiley face', can be interpreted in at least three ways: as a 'criticism' ('the system should have spotted an incorrect file'), as 'cheating' ('the system believed that work was submitted on time') or even as a 'insight into future design changes' ('the system doesn't check to see if a file is empty—they could do that in the next version!'). However, each of these adds connotation to what it represents, and the most valid interpretation should be taken from the actual accounts, and the alternative ways of expressing related ideas in similar accounts. Up to the end of phase 2 this is unproblematic. Any issues of concern here would be common to any other approach to QA.

On a larger scale, where a larger number of cases have been collected, some tests have been conducted to demonstrate multi-coder reliability. However, these are to predetermined coding frameworks, at the end of this phase, and they do not test whether different phenomenographers would produce the same coding structures when presented with the same set of accounts (Sandberg, 1997, p. 205)—that is, before phase 1.

Phase 3 should be more difficult to justify: by the end of this phase the outcome space should be independent of the original accounts, and so the meaning of any terms must become ‘fixed’ by other parts of the process—but must only be established from patterns within the data. When this phase of the evaluation study was reviewed during the teachback process, some anomalies arose over the concepts of complexity and inclusiveness within the final model. Two, in particular, suggested that additional value systems were influencing the final model, rather than the data itself.

The first anomaly arises from the suggestion that increasing complexity should reflect a higher level of understanding. This problem arose when we considered a more abstract coding of Table 2 (see Table 4). In this we have used abstract letters to ensure that the complexity of what is written is considered. When the code is stripped of any additional meaning, it is difficult to argue that the sequence that was produced in this phase in the evaluation study is justified. The comment at level 2 seems more complex than any of the others, and those at level 3 and 4 appear equally complex.

On reflection, the justification for this final sequence depended on the intended use of the tool as understood by the researcher conducting the analysis. As a tool for collaborative learning, the inclusion of ‘B’ (‘cheating’) has negative connotations, while the only comment that might even suggest an awareness of this possibility is ‘A’ (off-site). Although the example seems trivial, if it exists at this level then it is certain to be harder to detect and justify on a more complex example. The use of an external value system, as occurred implicitly in this case, can always be justified as part of the analysis, if it is acknowledged and independently justified. Once the possibility is raised that a value system has been introduced in this way, it might offer a different motivation for treating exchange of ideas above exchange of files (as in Table 3). There is a risk, however, that value systems are introduced unwittingly. If the best ways of understanding a phenomenon matches the ‘authorized’ view (Marton, 1981, p. 184), there is a risk that the researcher finds what they expect to find. This problem does not necessarily imply that the outcomes are dependent on the researcher, nor invalidate the analysis if such ‘values’ can be justified by the context of the research. However, it would seem essential to recognize that value systems that are external to the data are needed to convert simple complexity into levels of understanding. Neither should the justification be *post hoc*—that only informs you about the researcher’s values, or why certain ‘authorized’ versions are so readily found in educational studies.

Table 4. Table 2 redrawn with abstract codes

	Submitting work: loading a document	Aspect: features
4	[Loading: off-site]	X: A
3	[Loading: confirmation]	X: C
2	[Loading: confirmation and empty]	X: B and C
1	[Loading]	X

The second anomaly appeared to be the dependence on the use of presumed inclusiveness. The comments on sharing from the evaluation study provide the most evident example, although this was not needed at the levels above this—all the feedback comments occurred in accounts that also included comments on sharing. However, if the principle is over-used, there is a risk that an implicit decision is taken to place a comment at an ‘authorized’ level and then justify the additional details as a *post-hoc* rationalization. This would also become harder to notice in larger studies with more data and ‘discovery’ as the only process of development.

There remains the possibility that the method of data collection that was used in the evaluation study limited the complexity within the accounts. A request to produce more detailed, multi-layered accounts offers no direct benefit to any subject in doing so. However, if this is considered to be a possibility, then so is the converse: that a semi-structured approach generates accounts that are deeper (Webb, 1997). If this were so, then ‘natural’ complexity could be created by a willingness to conform to the interview setting (see Säljö, 1997), rather than being a simple reflection of the individual’s own awareness.

The generalizability of the outcomes

Phenomenography makes specific claims regarding the models that are created: once a model is discovered, then the understanding of every suitable subject of the same phenomenon must fit within the model. This claim is far stronger than most approaches to QA, even if the scope of the terms ‘suitable’ and ‘phenomenon’ might always allow counter-examples to be excluded as unsuitable or a different phenomenon.

Where other approaches produce a similar super-index as the endpoint of the research, they limit its meaning by the way it was produced. It is no more than a multi-dimensional index to the original accounts (for example, Säljö, 1997). Other approaches would require a test for validity, either at this stage, or after building a more complex model (e.g. ‘theoretical saturation’ in GT; Glaser & Strauss, 1967). Without any further test, and without any clear indication that the outcome space is independent of both the sample used and of the researcher, an ‘in principle’ argument seems unlikely.

From an ‘in practice’ perspective, the evaluation study is of limited assistance here. Students had to have studied on the module for a sufficiently long period of time to collect any relevant accounts, and beyond that point they were too committed to other assessments to allow a suitable empirical test; that is, to collect further cases and demonstrate that they do conform to the model. However, the literature is more useful here. Examples have been discovered where the categories of description have had to be modified, even with examples that were central to the original work in education. Following studies in other cultures in the early 1990s (Marton & Booth, 1997, pp. 39–45) it became clear that concepts that are central to the early work in phenomenography, such as ‘rote learning’ and deep and shallow learning, have had to be revised in specific subjects. This problem was ‘resolved’ by refining some critical terms, but the need to do this suggests that other issues would need empirical testing.

This suggests that the models are valid until they are shown to be invalid. This may be a problem that is shared with scientific models of knowledge, but the evidence for generalizability, in the case of phenomenography, seems far less than claimed.

The scope of phenomenography

The three issues discussed so far (collection of data, independence of the outcomes from the researcher and generalizability of the outcomes) all arose from the evaluation study, but the process study itself raised issues. This suggests that there is a limit to what can be observed within a phenomenographic study.

The concern arises from the timelessness of the phenomenographic research model. Data collection captures single 'snapshot' images for each individual and accepts that this is just one account from the wider set of accounts they could have produced. Since there is no attempt to 'assess' each student, there need be no concern whether this account is typical for this student or not. Phenomenography focuses only on the 'experience per se'—a current, individual explanation, without concern to understand previous experience as a formative process. Phenomenography describes what a deeper understanding might be, and the variation that might be experienced (Bowden & Marton, 1998), but not how this transfers to learning.

A second concern, which we characterize as collaborative accounting, follows on from this and reflects one of the distinctions between phenomenography and phenomenology. In the evaluation study, as elsewhere, the subjects will have developed their own awareness of this phenomenon and, on a number of occasions, discussed (accounted for) these experiences with other individuals, including members of their team. This opens up the possibility that the accounts that are collected at a later stage for analysis reflect previous patterns of accounting and provide an equally plausible justification for the limited number of 'possible experiences' that are found in this study, and indeed in any phenomenographic study. A similar effect would occur if the users are instructed or informed as to how they should use a system (as noted in a study by Orlikowski, 1992), and reflect this back as 'the right answer'. If we even allow the possibility these effects could influence the accounts that are produced, then the data collected are no longer objective (Bainbridge, 1981; Edwards & Mercer, 1987; Säljö, 1996).

Conclusions

None of these four issues are fatal to 'pure' phenomenography as an approach to QA, but each of them (data collection, independence of the results from the researcher, status of the results and scope of what is researched) provides a challenge.

In data collection, there is a thin line to be drawn between a method that will tease out an account of an experience, without the process corrupting or distorting the data, and removing the 'objectivity' that is claimed.

The independence of the results from the research can only be established if the researcher acknowledges that the analysis cannot be conducted without recognition

that ‘their’ values influence the coding that is used; that is, that independence requires both self-awareness and management.

The claim that the outcome space must limit any future descriptions seems unproven and may be unnecessary. If the impact of research in ICTE is to change future experiences, then it is difficult to see why we need to generalize across systems that cease to exist. However, without this claim, the outcome space of a phenomenographic study could be reduced to a useful cross-index. In that case, it could be argued, the outcome space has no special properties and ‘pure phenomenography’ offers nothing that is new.

Finally, the limitations on scope would appear impossible to address within phenomenography. Indeed, how can the development of understanding within an inherently social educational environment be researched if the model of knowledge can only capture accounts of isolated experiences, no matter how regulated these may be?

Acknowledgements

The researchers would like to acknowledge the input that was made to this project by Dr Chris Hutchison, at Kingston University, and all other members of the NetPro project. The authors are also grateful to Jane Lindsay, Hilary Tompsett and Dr Linda Burke for their comments on the final stages of this paper.

Note

1. Innovative integration of problem-based learning and network technology funded under the Leonardo programme to design, develop, test and disseminate network-based project learning approaches in engineering education.

References

- Alsop, G. & Tompsett, C. (2002) Grounded theory as an approach to studying students’ use of learning management systems, *ALT-J*, 10, 63–76.
- Ashmore, M. (1989) *The reflexivity thesis* (Chicago, The Chicago University Press).
- Bainbridge, L. (1981) Verbal reports as evidence for the process operator’s knowledge, in: E. H. Mamdani & B. R. Gaines (Eds) *Fuzzy reasoning and its applications* (London, Academic Press).
- Booth, S. (1997) On phenomenography, learning and teaching, *Higher Education Research and Development*, 16, 135–157.
- Bowden, J. & Marton, F. (1998) *The university of learning: beyond quality and competence* (London, Routledge Falmer).
- Dahlgren, L. O. & Marton, F. (1978) Students’ conceptions of subject matter: an aspect of learning and teaching in higher education, *Studies in Higher Education*, 3, 25–35.
- Edwards, D. & Mercer, N. (1987) *Common knowledge: the development of understanding in the classroom* (London, Routledge).
- Entwistle, N. (1997) Phenomenography in higher education, *Higher Education Research and Development*, 16, 127–134.
- Glaser, B. G. & Strauss, A. L. (1967) *The discovery of grounded theory: strategies for qualitative research* (London, Weidenfeld and Nicholson).

- Hasselgren, B. & Beach, D. (1997) Phenomenography—a 'good-for-nothing brother' of phenomenology? Outline of an analysis, *Higher Education Research and Development*, 16, 191–202.
- Irvine, S. (2002) Parent conceptions of their role in early childhood education and care public policy and service development, paper presented at the *Current Issues in Phenomenography Symposium*, Canberra.
- Kidd, A. L. (Ed.) (1987) *Knowledge acquisition for expert systems* (London, Plenum).
- Marton, F. (1981) Phenomenography—describing conceptions of the world around us, *Instructional Science*, 10, 177–200.
- Marton, F. (1986) Phenomenography—a research approach to investigating different understandings of reality, *Journal of Thought*, 21, 28–49.
- Marton, F. (1994) Phenomenography, in: T. Husen & T. N. Postlethwaite (Eds) *The international encyclopedia of education* (2nd edn) (Oxford, Pergamon).
- Marton, F. & Booth, S. (1997) *Learning and awareness* (Mahwah, NJ, Lawrence Erlbaum Associates).
- Marton, F. & Pang, M. F. (2003) Beyond 'lesson study': comparing two ways of facilitating the grasp of some economic concepts, *Instructional Science*, 31, 175–194.
- Marton, F. & Säljö, R. (1976a) On qualitative differences in learning: I. outcome and process, *British Journal of Educational Psychology*, 46, 4–11.
- Marton, F. & Säljö, R. (1976b) On qualitative differences in learning: II. outcome as a function of the learner's perception of the task, *British Journal of Educational Psychology*, 46, 115–127.
- Marton, F. & Säljö, R. (1984) Approaches to Learning, in: F. Marton, D. Hounsell & N. Entwistle (Eds) *The experience of learning* (Edinburgh, Scottish Academic Press).
- Mugler, F. & Landbeck, R. (1997) Learning in the South Pacific and phenomenography across cultures, *Higher Education Research and Development*, 16, 191–202.
- Orlikowski, W. (1992) Learning from notes: organizational issues in groupware implementation. In proceedings of the 1992 ACM conference on Computer-Supported Cooperative Work, 1–4 November, Toronto, Ontario, Canada. (New York, ACM Press), 362–369. Available online at: <http://doi.acm.org/10.1145/143457.143549> (accessed 14 July 2006).
- Pask, G. (1977) *Conversation, cognition and learning* (New York, Elsevier).
- Ramsden, P., Masters, G., Stephanou, A., Walsh, E., Marint, E., Laurillard, D., et al. (1993) Phenomenographic research and the measurement of understanding: an investigation, *International Journal of Educational Research*, 19, 301–316.
- Sandberg, J. (1997) Are phenomenographic results reliable?, *Higher Education Research and Development*, 16, 173–190.
- Säljö, R. (1996) Minding action—conceiving of the world versus participating in cultural practices, in: G. Dall'alba & B. Hasselgren (Eds) *Reflections on phenomenography—toward a methodology?* (Göteborg, Acta Universitatis Gothoburgensis).
- Säljö, R. (1997) Talk as data and practice—a critical look at phenomenographic inquiry and the appeal to experience, *Higher Education Research and Development*, 16, 173–190.
- Svensson, L. (1997) Theoretical foundations of phenomenography, *Higher Education Research and Development*, 16, 159–171.
- Webb, G. (1997) Deconstructing deep and surface: towards a critique of phenomenography, *Higher Education*, 33, 195–212.
- Widang, I. & Fridlund, B. (2003) Self-respect, dignity and confidence: conceptions of integrity among male patients, *Journal of Advanced Nursing*, 42, 47–56.