Investigating Incorrect Understandings of CS Concepts (A Discussion Proposal)

Juha Sorva
jsorva@cs.hut.fi

1 Background: Phenomenography in My Current Research

The phenomenographic research approach is helpful for investigating students' understandings of concepts they are expected to learn. As a result, phenomenography is of great interest for me, as I am currently exploring beginner programmers' understandings of object-oriented program execution and related concepts. My work in the spring and summer of 2008 revolves around an interview study of students' understandings of the concept of variable. The current study ties in with my prior phenomenographic study on students understandings of storing objects in the computer [5]. In my current study, as in the previous one, I am interested not only in discovering different kinds of correct understandings that students have of a concept but also different incorrect understandings. There are different approaches to dealing with correctness and incorrectness of understandings in a phenomenographic study. This paper enumerates three different approaches to dealing with incorrectness and proposes a discussion on the issue.

2 Introduction

In computer science, there are concepts that are clearly defined and that serve as well-defined intended learning outcomes for students of computer science. Unlike in some other disciplines, it is meaningful for computer science instructors to assess the correctness of their students' understandings of such concepts. A qualitative study of student understandings of a well-defined computer science concept is likely to discover understandings that conform with the intended learning outcome (and can be said to be *correct*) as well as understandings that are wholly or partially unwanted (*wholly incorrect* or *partially incorrect*).

Delimiting what constitutes a phenomenon in a phenomenographic study is a researcher-driven process that is affected by the researcher's own understanding of the phenomenon as they choose which parts of their data are related and which are not. The researcher must be careful so as not to create over- nor under-inclusive outcome spaces. If the intention is to examine understandings of a single phenomenon, one must be careful not to accidentally treat multiple different phenomena as a single one. For a phenomenographer investigating students' understandings of a computer science concept, explicitly assessing the correctness of understandings during phenomenographic analysis is potentially helpful in the key task of delimiting the phenomenon.

3 Three Ways of Dealing with Incorrectness

Below, I describe three approaches to dealing with incorrectness of understandings during phenomenographic analysis of people's understandings of a concept.

1. The Equal Approach: Include in the outcome all categories of understandings deemed to be relevant. That is, include the entire set of relevant understandings expressed by the cohort, treating correct and incorrect understandings equally. Do not explicitly draw on a definition of correctness when delimiting the phenomenon and analyzing the data. Leave making judgements about correctness for later, possibly to third parties reading the results you report.

Many phenomenographic studies, even those related to the 'harder sciences', do not problematize the correctness of understandings expressed in the data. Assessing the correctness may be an issue that is deemed unnecessary or irrelevant for the task of creating an outcome space, as all categories of understanding are deemed equally valid. For instance, Adawi and Ingerman [2] paraphrase work by Adawi and Linder [1] on lay adults understandings of the concepts of heat and temperature. In addition to more orthodox understandings such as "Heat is the internal kinetic energy of the atoms or molecules in an object.", the resulting outcome space also contains understandings such as "Heat is a substance consisting of some kind of particles.", which is incorrect within the paradigm of modern physics. Judgements about correctness do not play an explicit part in analyzing the data and forming the outcome space.

When using the equal approach, delimiting the phenomenon is challenging when partially or wholly incorrect understandings are expressed by a cohort. One needs to be careful in justifying that the various categories describe a single phenomenon being investigated, not multiple, possibly imaginary phenomena. What are the grounds for including something in the data in the analysis, but discarding something else from the outcome space? Let us consider a study that investigates CS students' understandings of objects. One interviewee may use the CS term 'object' to talk about objects, another about object-valued variables, while a third individual uses it to talk about classes, and a fourth one about a concept that is a curious merge between the textbook definitions of 'object', 'variable' and 'file'. It is the researcher's job to discover which concept each individual talks about and therefore what parts of data are relevant and will be used to form the outcome space. If the intention is to study understandings of the phenomenon (not the word) 'object' in programming, the researcher may want to draw on the correct definition of the term to help himself decide which utterances are relevant in terms of his research goal. This observation leads to the two alternative approaches described below.

2. The Clear Cut Approach: Include in the outcome space only those aspects of understandings which you deem to be correct as defined by the scientific paradigm or technical definition that provides the intended learning outcome. Discard all aspects of understandings that you deem incorrect.

The clear cut approach seems to be relatively common within computer science education research, even if the matter of correctness is not always explicitly treated. (See e.g. Eckerdal and Thuné's work on understandings of objects and classes [4] and Berglund's work on understandings of network protocols [3].) This is not surprising since many such studies are interested in correct but partial understandings which describe 'the real phenomenon', not in discovering a wider range of possibly incorrect categories of understandings. In studies of this type, "concentrating only on the correct understandings is both enough and convenient [...] and allows the researcher [...] to produce neater hierarchies that are arguably easier to report, understand, and make use of" [5]. Relying on a definition of correctness also helps with delimiting the phenomenon, and making sure one stays focused on a single phenomenon.

A third approach attempts a compromise between the two described above.

3. The Anchored Approach: Include correct aspects of understandings in the outcome space, as in the clean cut approach above. Additionally, include such incorrect understandings that extend the correct understandings in an incorrect way and are

therefore partially incorrect. However, discard understandings that do not extend the correct understandings, as they are deemed wholly incorrect and outside the phenomenon being investigated.

An example this approach is my own work on introductory programming students' understandings of storing objects [5]. Like the equal approach, the anchored approach allows studying relevant incorrect understandings that a cohort expresses, inasmuch as they are partially incorrect 'over-extensions' of correct understandings. The correct categories of the outcome space serve to 'anchor' the partially incorrect ones. As with the clear cut approach, using the anchored approach involves the researcher in judgements about the correctness of understandings during phenomenographic analysis. This can help delimit the phenomenon neatly and make sure a single phenomenon is being focused on, thereby improving the trustworthiness of the research.

4 Questions for Discussion

- Is the above categorization of approaches helpful?
- What are pros and cons of the various approaches?
- Can delimiting a phenomenon in terms of correct understandings (as in the clear cut approach and the anchored approach) help in establishing trustworthiness of research, as suggested above?
- What might be the best way, in practise, to go about implementing the anchored approach? Perhaps an initial round of analysis focusing on correct categories of description, followed by another round that explores possible over-extensions of the correct categories?

References

[1] T. Adawi and C. Linder. What's hot and what's not: A phenomenographic study of lay adults' conceptions of heat and temperature. In *The 11th biennial EARLI conference*, 2005.

[2] Tom W. Adawi and Åke Ingerman. Phenomenography for physicists. In Michael F. Christie, editor, *Shifting Perspectives in Engineering Education*, pages 263–272. Chalmers Strategic Effort on Learning and Teaching, 2006.

[3] Anders Berglund. Learning computer systems in a distributed project course: The what, why, how and where, 2005.

[4] Anna Eckerdal and Michael Thuné. Novice Java programmers' conceptions of "object" and "class", and variation theory. *SIGCSE Bulletin*, 37(3):89–93, 2005.

[5] Juha Sorva. Students' understandings of storing objects. In *Proceedings of the Seventh Baltic Sea Conference on Computing Education Research (Koli Calling 2007)*, pages 127–136, 2008.