A review of conceptual change research in science education

Una revisión de las investigaciones sobre el cambio conceptual en la educación científica

CARLOS SOTO-LOMBANA¹, JOSÉ OTERO², VICENTE SANJOSÉ³

¹ Departamento de Educación Avanzada, Universidad de Antioquia, Calle 67 # 53-108. Medellín-Antioquia, Colombia
² Universidad de Alcalá de Henares, Spain. ³ Universidad de Valencia, Spain.
csoto@ayura.udea.edu.co

Abstract

This article describes research done on conceptual change as published in four of the better-known science education journals during a 20-year period, from 1981 to 2001. The present review was focused on identifying three characteristics of that research: a) temporal distribution, b) research topics, and c) methodology: experimental designs and validity. The results show that much research work was carried out, although essential elements of Conceptual Change remained unclear. With respect to the methodology, we have evidenced scarce replication of previous studies, and an important number of studies that fall short of desirable validity levels.

Key words: Conceptual Change, Learning, Review, Science Education.

INTRODUCTION

Research on science learning has repeatedly shown that students have conceptions associated to explanations of natural phenomena that clash with accepted scientific ideas. These misconceptions are remarkably similar among students from different cultures. The attempts in modifying these misconceptions evidenced the difficulties associated with this task. The entire modification effort was termed "Conceptual Change" (CC), and became a topic of research in itself. Studies on CC deal with how to change the non-scientific belief systems used by students to explain nature and how to turn these into scientifically valid knowledge. For this, attention is paid to the interaction between the learner's naive knowledge based on his/her everyday experience and new knowledge acquired through instruction.

CC studies originated in the early 80's as an outcome of research on alternative conceptions. Hewson (1981, 1982) and especially the Cornell group, composed by Posner, Strike, Hewson and Gertzog (1982) laid the cornerstone for research in this field. Thus, Posner et. al. (1982) used Kuhn and Lakatos's philosophical ideas about change in scientific theories as an analogy for conceptual change in an individual.

The seminal ideas (Duschl & Hamilton 1992) about CC, presented in Posner et. al.'s (1982) paper, were refined in the following years (Strike & Posner 1985; Hewson & Thorley 1989; Hewson & Hewson 1992), and even reformulated one decade after the initial article was published (Strike & Posner 1992). The original work was grounded in the philosophy of science. It gave birth to a research program (Kelly 1997) that analyzed changes in scientific theories in order to identify factors that facilitate or hinder the process of acquisition of scientific concepts by the individual.

Since then CC has evolved into an important research area dealing with teaching and learning science throughout the past two decades:

- Conceptual change undoubtedly has been the most powerful frame for research on teaching and learning science for the past 25 years. There is no doubt that the present state in the research domain allows to understand teaching and learning processes much better than in the beginning of the 1980s.(Duit 2002, p. 5).

- In view of the above, a revision focused on the state of this research program and encompassing its 20 years of existence seems justified.

In continuation we shall list selection criteria used for choosing the sample of analyzed papers on CC in the area of science education. This is followed by the description of the analyzed variables and the outcome of the analysis. Finally, we will discuss the most important results and their implications in the CC field.

METHOD

Selection of the articles

We have chosen International ERIC Database as our main source of information. The search was limited to documents published in Journal or Review, during 1980-2001 interval, and we used the term Conceptual Change as Identifier, with Science Education as keyword. Thus we were able to identify 117 articles published in 20 different journals. It so happened that 4 influential journals covering the area of Science Education such as Science Education, International Journal of Science Education (formerly European Journal of Science Education), Journal of Research in Science Teaching and Research in Science Education, contained 78.5% of all the compiled information. We resorted to three criteria while making this second selection:

- The article's topic had to deal with CC. By analyzing the article's title and its abstract we were able to identify the topic.
- The article should include empirical results involving subjects, analysis of the results, interpretation as well as conclusions.
- The research must have been carried out in the field of science teaching or learning, within a formal educational setting, at any education level, or under laboratory conditions.

This second search resulted in the selection of 59 articles that fulfilled the previously listed conditions (see Appendix).

RESULTS

Temporal distributions of research work

Figure 1 presents the distribution of articles during the studied period. It shows that research on CC has taken place in the last decade and that 49.2% of the articles were published in the last five years.

Research topics

Education level

Figure 2 shows research according to subjects' education level. The greatest part of research was done on subjects from high schools.

Scientific topics

Different scientific topics have received unequal attention in research on...
Evaluating Instructional Strategies by means of sampled articles (44.1% of all the studies), and its importance increased in four cited conditions for CC. They summed up to a substantial part of the learning processes, conceptual ecology or the relative importance of the way the subjects’ conceptions are restructured. It also involves studies on metacognitive aspects of CC. They analyze key elements of CC such as the tests of CC models”. These studies go beyond instructional efficiency or et. al. (1982) four conditions for CC. Fifty percent of the studies from this propose tools for CC. Consequently; there is usually no reference to Posner’s less emphasis on instructional strategies because their aim was not to students’ alternative conceptions. Nonetheless, the examined studies placed natural selection, or chemical equilibrium, and how the corresponding students’ alternative conceptions; they focus on such topics as energy, change”. The purpose of these articles is to describe the evolution of half of that period.

The most frequent setting for carrying out the studies was science lectures: 88.1% of the studies have been carried out within this scenario. The most frequent setting for carrying out the studies was science lectures: 88.1% of the studies have been carried out within this scenario. The purpose of these articles is to describe the evolution of half of that period.

The third research area involved studies that focused on “Analysis and tests of CC models”. These studies go beyond instructional efficiency or the characteristics of a particular student’s conceptions; as in the previously cited research, and focus instead on the ontological, epistemological and metacognitive aspects of CC. They analyze key elements of CC such as the way the subjects’ conceptions are restructured. It also involves studies on learning processes, conceptual ecology or the relative importance of the four cited conditions for CC. They summed up to a substantial part of the sampled articles (44.1% of all the studies), and its importance increased in the 90s, since 76.9% of these studies appeared between 1996 and 2000.

Table 2 shows how many studies belonged to each of the three areas.

<table>
<thead>
<tr>
<th>Research Areas</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluating Instructional Strategies by means of Conceptual Change</td>
<td></td>
</tr>
<tr>
<td>Cooperative learning</td>
<td>6</td>
</tr>
<tr>
<td>Use of computers</td>
<td>4</td>
</tr>
<tr>
<td>Use of analogies and mental models</td>
<td>6</td>
</tr>
<tr>
<td>Use of historical arguments</td>
<td>1</td>
</tr>
<tr>
<td>Use of writing</td>
<td>1</td>
</tr>
<tr>
<td>Use of refutational text</td>
<td>1</td>
</tr>
<tr>
<td>Use of conceptual substitution</td>
<td>1</td>
</tr>
<tr>
<td>Mixed language strategy</td>
<td>1</td>
</tr>
<tr>
<td>Students’ Conceptions and Conceptual Change</td>
<td></td>
</tr>
<tr>
<td>Physic targets</td>
<td>5</td>
</tr>
<tr>
<td>Chemistry targets</td>
<td>5</td>
</tr>
<tr>
<td>Biological targets</td>
<td>2</td>
</tr>
<tr>
<td>Analysis and test of CC models</td>
<td></td>
</tr>
<tr>
<td>Nature and complexity of Conceptual Change</td>
<td>12</td>
</tr>
<tr>
<td>Metacognitive aspects</td>
<td>4</td>
</tr>
<tr>
<td>Ontological and epistemological issues</td>
<td>6</td>
</tr>
<tr>
<td>Conceptual ecology</td>
<td>2</td>
</tr>
<tr>
<td>Motivation issues</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2. Number of studies for each research type

<table>
<thead>
<tr>
<th>Article cited</th>
<th>% of articles citing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Champagne, Gunstone &amp; Klofer (1985)</td>
<td>28.8</td>
</tr>
<tr>
<td>Hewson &amp; Thorley (1989)</td>
<td>23.7</td>
</tr>
<tr>
<td>Pintrich, Marx &amp; Boyle (1993)</td>
<td>23.7</td>
</tr>
<tr>
<td>Strike &amp; Posner (1992)</td>
<td>22.0</td>
</tr>
<tr>
<td>Osborne &amp; Freyberg (1985)</td>
<td>18.6</td>
</tr>
</tbody>
</table>

Table 3. Most frequently cited articles

With respect to the most frequently cited authors (15 or more quotations), Driver received the bulk of quotations (cited in 62.7% of sampled articles), P. Hewson (40.0%) and Osborne (32.2%) came second and third, respectively. It should be pointed out that these quotations corresponded to articles’ content, not necessarily dealing with CC topic. Also, the articles were authored by these researchers alone or together with other coauthors.

Methodological characteristics

The sampling was done on articles containing different methodological approaches, including experimental, quasi-experimental, and naturalistic methods. Several characteristics have been sought in all articles: sample’s cohort size, duration, setting and type (quantitative vs. qualitative). In addition, internal and external validity have been appraised using different criteria for quantitative and qualitative studies.

Sample, duration, setting and type

With respect to sample, the studies ranged from 1 subject only (# 46 and # 57) to 310 (# 18). Social and cultural variables have not been explicitly considered in any of the studies. Duration of the study was also variable, from 20 minutes (# 16) to two school-years projects (# 57).

The most frequent setting for carrying out the studies was science lectures: 88.1% of the studies have been carried out within this scenario. Of course, such a complex setting makes it difficult to use experimental designs and to insure the internal validity of the quantitative studies. This issue is discussed farther on. The remaining 11.9% of the studies were done under more controlled, laboratory conditions.

With respect to the quantitative/qualitative ratio, there was higher percentage of qualitative studies, 55.9%, versus quantitative counterparts, 44.1%. The publishing frequency was also uneven: quantitative studies predominated during the first half of the 90’s, while qualitative studies were published more often during the second half of the decade. See figure 3.

![Figure 3. Temporal distribution of quantitative vs. qualitative articles](image)

Internal and external validity

Internal validity in quantitative studies was appraised according to criteria based on Campbell and Stanley’s (1963) classical study. In particular, we examined sampled articles while looking for ‘the explicit mention in the text of the article’ of: a) the existence of a control group, b) random assignment of subjects, c) existence of pretest and posttest, and d) other threats to internal validity. Among the latter we have included maturation, experi-
mental mortality, and testing. In view of the all above, internal validity was graded low, medium or high.

Credibility is a concept similar to internal validity when dealing with qualitative studies. Thus, credibility was appraised according to the criteria developed by Lincoln and Guba (1985). These include: a) persistent observation, i.e., the observer should stay in the field long enough to assure data's consistency, b) triangulation of observers, i.e., presence of two or more observers, and triangulation of measurements, i.e., using a variety of measuring methods, c) peers' critical judgment to reduce researcher's biases, d) use of reference materials, including documents, films and audio tapes that would allow analyses and realyses, e) verification of researchers' interpretations against the studied subjects.

Credibility judgments were made 'looking for explicit mention in the text of the article'. Persistent observation was always requested -for at least several days. Verifying through participants was discarded because of the difficulties to do this with children who participated in many studies. Consequently, credibility was scored depending on the number of observers and triangulation of measures. Secondly, when other threats existed such as lack of audiotape or video records, absence of explicit evaluation criteria for categorization or rating the subjects, and no-revision clues for supervision or replication of the methodological steps, credibility was decreased one degree.

Table 4 shows that 30.5% of the total number of studies had low internal validity or credibility. On the other hand, 30.7% of the quantitative studies attained high internal validity, and 39.4% of qualitative studies had high credibility.

<table>
<thead>
<tr>
<th>Internal Validity/ Credibility Score</th>
<th>Quantitative (%)</th>
<th>Qualitative (%)</th>
<th>Total (quantitative + qualitative) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>30.7</td>
<td>39.4</td>
<td>35.6</td>
</tr>
<tr>
<td>Medium</td>
<td>38.6</td>
<td>30.3</td>
<td>33.9</td>
</tr>
<tr>
<td>Low</td>
<td>30.7</td>
<td>30.3</td>
<td>30.5</td>
</tr>
<tr>
<td>Totals</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 4. Percentage of quantitative studies and qualitative studies according to internal validity score or credibility score

DISCUSSION

We were interested in getting answers to the three main questions posed at the beginning of this article. In continuation we shall deal with each one of them separately.

With respect to first question, we have found that international community interested in research on CC is composed of researchers from different countries and continents, mainly Australia, Europe and North America. But there is a clear dominance of USA researchers over Europeans and Australians.

Although the key papers on Conceptual Change were published at the beginning and halfway through the 80's, a substantial production of papers did not pick up until the decade of the 90's. This may be due to the fact that many researchers in science education were still working on the identification and description of misconceptions in science-an area that occupied much of research time on science education since the late 70's.

With respect to education level, the highest frequency of studies at the secondary level (high school) seems to reflect a research characteristic in the reviewed sample. The small number of studies at the tertiary level coincided with the general situation of science education studies concerning college students: traditionally less attention was paid to learning problems at this level as compared to the concerns of teaching adequate content.

Within this wide sample of studies we have noticed a predominance of topics on mechanics and biological evolution. The predominance of topics on mechanics may be partly explained by its immediacy, by its importance in Physics curricula, and by historical reasons: concepts in mechanics were the object of the first studies on students' alternative conceptions. Biological evolution forms an essential paradigm in Natural Sciences.

We have classified articles according to three areas of research. Articles in the first area "Instructional strategies for CC" were based on the assumption that instruction aimed at achieving CC should reflect the four conditions for change stated by the Cornell group in the 1982 paper. However Strike and Posner (1992) cautioned against a rigid translation of these ideas into instructional procedures.

The second area 'Students' conceptions and conceptual change' represents continuity with the studies on misconceptions carried out in the past.

The third area "Analysis and tests of CC models" gained importance in recent years as mentioned above.

Afterwards, we proceeded to analyze links within the sample and outside works to find out about the theoretical and empirical background of research on CC. Unsurprisingly, the article by Posner et. al. (1982) was the most frequently cited in the sampled articles and constitutes the most important theoretical foundation for research on CC. Likewise, researchers belonging to Cornell Group produced several of the most frequently cited papers: Posner, Strike & Hewson, respectively authored two papers among the six of the most widely cited.

The ideas presented in the seminal paper by Posner et. al. (1982) were explained and extended in the 80's (Hewson, 1981; Strike & Posner, 1985; Hewson & Thorley, 1989) and later reformulated by Strike and Posner (1992). However, these additions and modifications were cited by less than 10% of the sampled articles. Since 74.6% of the articles in our sample were published after 1992, we concluded that researchers were not interested in these modifications. An alternative explanation for little impact of these later studies may be that they were not published in better-known journals. This assumption could be justified by the fact that another variant of the original ideas summed up in a book by Hewson and Thorley (1989) originally published in International Journal of Science Education, is one of the six most frequently cited articles.

With respect to the methodological quality of research, we have found that only 30.7% of the quantitative studies had high internal validity, and only 39.4% of qualitative studies shared the same credibility. Threats to internal validity are associated with nonrandom convenience samples without a pretest, effects of instrumentation on the measured outcomes, and poor description of procedures should a replication be attempted. This last problem was especially true for qualitative studies.

There were frequent problems related to the validity of measuring instruments as well. Thus, in 31.0% of the quantitative studies we found an absence of explicit validation of the measuring instrument or/and scoring criteria. In 72.2% of qualitative studies there was an lack of criteria supporting a particular selection of information in students' cognitive structures, and/or explicit criteria for categorizing students' responses.

As previously shown, methodological approaches experienced a shift from predominantly quantitative to qualitative studies. However, the latter evidenced methodological problems too, basically related to the absence of precise safeguards such as triangulation, inter-judge agreement, and appropriate records for controlling and reducing researchers' biases.

CONCLUSIONS AND RECOMMENDATIONS

In retrospective, our review presents some hints suggesting that research on CC is still far from maturity. We have found serious deficiencies in the methodological aspect of that research, such as a lack of tradition in replicating previous studies, compounded by the high number of studies below a desirable level of validity.

Many of the points described in this review require the concurrence of the international community in order to establish, in a concerted way, the criteria and recommendations on which a progressive Research Program could be based, with agreed upon guidelines on the basic theory, the questions, the hypothesis, the methodological procedures and the plan of action for the research. An effort of this nature, as well as being important, would make a substantial advance possible in the research in this field and, at the same time, have a positive effect on the field of Science Education as a whole.

ACKNOWLEDGEMENTS

The research was possible thanks to the grant secured by the third author of this article, who was sponsored by the Universidad de Antioquia, Colombia-South America, and by Grant BS--02002-02736, of the Ministry of Science and Technology, Spain, secured by the first and third authors. We thank Jerry Keller, from the Universidad de Alcalá, for the linguistic revision of the manuscript.

BIBLIOGRAPHY


