Effectiveness of conceptual change texts: A meta analysis

Fulya Öner Armağan¹, Melike Özer Keskin² And Beril Salman Akın³
1 Department of Mathematics and Science Education, Erciyes University, Kayseri, Turkey
2, Department of Mathematics and Science Education, Gazi University, Ankara,Turkey
3 Department of Mathematics and Science Education, Gazi University, Ankara,Turkey
For correspondence: onerf@erciyes.edu.tr

Abstract
The purpose of this study was to determine the overall effectiveness of conceptual change texts (CCTs) on academic achievement and to find out if effectiveness was related to some characteristics of the study. It followed up a Meta-analysis research approach. 42 published and unpublished studies, published between 1995 and 2010, and 42 experiment groups’ effect sizes were tested in this study. The overall effect size for CCTs was calculated as 1.18. This is a large effect size according to Cohen’s criteria. It was found that CCTs have been quite successful in promoting the students’ academic success. However results showed that there was no publication bias in respect to publication status. Moreover, no statistically significant difference was found on the nature of assessment instrument, study origin, subject matter, type of instruction, school level, sample size, type of instrument, publication date and duration of treatment. Moreover statistically significant difference was found for instruction effect. The effects of the research results are discussed.

Keywords: Conceptual change, conceptual change text, meta-analysis

Introduction
There have been many studies concerning students’ conceptual understanding since 1970’s. All these studies’ main concern was the importance of on students’ prior ideas (Pintrich, Marx ve Boyle, 1993;Scott, Asoko & Driver, 1992; Smith, diSessa & Rochelle, 1993; Chambers & Andre, 1995; Beeth, 1998a). Students come to the classroom with some alternative ideas which contradict scientific ideas (Posner, Strike, Hewson & Gertzog, 1982; Roth 1985; Gunstone & Nortfield, 1992; Vosniadou & Ioannides, 1998). Since conceptions have an active role on teaching and learning, the studies on students’ and teachers’ conceptions are one of the most important areas of science education research (Pines and West, 1986; Duit & Treagust, 2003). Students’ prior knowledge has an effect on their learning (Pines and West 1986). Hence, determining students’ prior knowledge (Beeth, 1998) and fostering a learning environment is quite important (Clough & Driver, 1986; Carey, 2000).

Many studies concerning alternative concepts have shown that traditional instruction has not fit concept learning and is not effective for removing alternative concepts (Pines and West, 1986; Dykstra, Boyle and Monarch, 1992). Therefore teachers need to take into some alternative strategies for fostering learning outcomes. Conceptual change is one of the effective strategies in science learning (Hewson, 1981; Gunstone & Nortfield, 1992; Chambers and Andre, 1997; Treagust and Duit, 2008).

¹ This study was presented as an oral presentation at the World Conference on Learning, Teaching and Administration (WCLTA 2010), 29-31 November 2010.
This study has been derived from the first author’s doctoral dissertation completed at Gazi University Graduate Institute of Educational Sciences.
Moreover this strategy is quite effective in reducing alternative conceptions (Hewson & Hewson 1983; Wang & Andre, 1991; Smith, Blakeslee & Anderson, 1993).

There are numerous strategies using conceptual change approach. One of them is conceptual change text (CCT) suggested by Roth (1985). Many studies concerning CCTs in science education have shown that this approach is quite effective on increasing students’ academic success (Chambers & Andre, 1997; Wang & Andre, 1991).

**Purpose of the Study**

The purpose of this study was to determine the overall effectiveness of CCTs on academic achievement and to find out if effectiveness was related to some characteristics of the study. Moreover, in present study, the data were examined to investigate the influence of the type of publication status, publication origin, publication date, subject matter, school level, sample size, length of treatment, nature of assessment instrument, type of instrument, mode of instruction and effects of intervenor on the overall effect size. This meta analysis study was expected to contribute in the literature as a comprehensive and updated analysis on the effectiveness of CCTs in science education.

**Method**

This study followed up a Meta analysis research approach. The meta analysis refers to an analysis in order to obtain an overall result by combining the results obtained from different studies (Hunter ve Schmidt, 1990). Meta analysis allows the combining of numerical results from a few or many studies (Rosenthal 1991). In meta-analysis, the data obtained in order to bring together the statistical data in different studies needs to be translated to the effect size, which is a common unit of a measurement. Effect size is a standard measure used in determining the strength of a study. Therefore, in this study effect sizes for studies and experiment groups were calculated using the formula recommended by Hedges and Olkin (1985). In this study only research studies that used a control group and provided adequate statistics for conversion to effect sizes were included. However effect size of each study was calculated and examined for outliers. Then present data analyzed through SPSS 18.1 and Meta Win 1.00 (Rosenberg, Adams & Gurevitch, 1997). The overall effect size of the all studies was calculated by Meta Win 1.00. In the study, a random effects model was chosen to calculate the mean effect size in the Meta Win 1.00.

**Data Collection and Analysis**

Firstly a comprehensive search was done selecting all of the related studies whether published or unpublished. When the studies were being searched during literature review, all studies were examined. Because the aim was to reach all related studies to CCT. To achieve publications in Turkey, Higher Education Council Dissertation Center, Ulakbim and Google Scholar Database was scanned. To achieve related studies, “kavramsal değişim metni”, “kavramsal değişim yaklaşımı” keywords were searched in databases. At the end of literature review, a master thesis, a doctoral dissertation and research paper related to CCT was found. The references of all studies were also examined to find relevant studies. All these studies were examined for selecting criteria and then selected studies was coded. Studies which have got insufficient information to calculate the effect sizes, connection to authors had done. The database was updated periodically.

The coding form was developed in order to increase the reliability in the coding of the suitability of the studies involved. Two researchers filled the coding form separately and then reached consensus on non-overlapping coding. To assess the inter-coder reliability, “agreement rate” (AR) was used. The formula for AR, as represented by Cooper & Hedges (1994) was used. An agreement rate of .85 or
greater was predetermined to be considered sufficient. An agreement rate of .99 was obtained for this study.

Selection of Studies

Depending on the inclusion criteria determined and on the availability of the studies, a total of 44 studies were included in the present study. The criteria for inclusion in this study was:

1. The studies on CCTs using pretest-posttest-control group design.
2. The studies examining the effect of CCTs on academic achievement of students
3. The studies including the data of mean scores, standard deviations and sample size for the experimental and control groups in order to calculate effect size.

Findings and Results

The overall effect size for CCTs was calculated as 1.20. This is a large effect size according to Cohen’s criteria. In present study, the data were examined to investigate the influence of the type of publication status, publication origin, publication date, subject matter, school level, sample size, length of treatment, nature of assessment instrument, type of instrument, mode of instruction and effects of intervener on the overall effect size. The results of this analysis are presented in Table 1 to table 11.

### Table 1. Effect sizes for publication status.

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect (Q_{bet})</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d</th>
<th>Homogeneity within Each class Q_{wi}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication status</td>
<td>0.460</td>
<td>40</td>
<td>1.0</td>
<td>0.838 - 1.352</td>
<td>40.372</td>
</tr>
<tr>
<td>Unpublished</td>
<td>15</td>
<td>1</td>
<td>1.0</td>
<td>0.838 - 1.352</td>
<td>40.372</td>
</tr>
<tr>
<td>Published</td>
<td>27</td>
<td>1.2</td>
<td>1.018 - 1.392</td>
<td>102.188</td>
<td></td>
</tr>
</tbody>
</table>

* p < .05,

The results of the analysis showed that there were no significant differences between the mean effect sizes of published and unpublished studies (Q_{bet} = 0.460, p=.497). The effect sizes for published and unpublished studies were nearly equal and they were $d=1.20$ and $d=1.09$ respectively. These findings represent that there was no publication bias in respect to publication status.

### Table 2. Effect sizes for study origin

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect (Q_{bet})</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d</th>
<th>Homogeneity within Each class Q_{wi}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Origin</td>
<td>3.778</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctoral dissertations</td>
<td>4</td>
<td>0.875</td>
<td>0.392 - 1.358</td>
<td>5.547</td>
<td></td>
</tr>
<tr>
<td>Master theses</td>
<td>9</td>
<td>1.070</td>
<td>0.740 - 1.401</td>
<td>23.204</td>
<td></td>
</tr>
<tr>
<td>Articles</td>
<td>27</td>
<td>1.205</td>
<td>1.020 - 1.390</td>
<td>102.188</td>
<td></td>
</tr>
<tr>
<td>Conference papers</td>
<td>2</td>
<td>1.651</td>
<td>0.960 - 2.342</td>
<td>1.207</td>
<td></td>
</tr>
</tbody>
</table>

* p < .05,
There were no statistically significant differences the mean effect sizes of among doctoral dissertations, master theses, articles and conference papers ($Q_{bet} = 3.778$, $p = .286$). Conference papers attained the highest effect size ($d = 1.651$). It was followed by articles and master theses.

Table 3. Effect sizes for subject matter

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect ($Q_{bet}$)</th>
<th>n</th>
<th>Mean d</th>
<th>$95%$ CI for d</th>
<th>Homogeneity within Each class $Q_{wi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject Matter</strong></td>
<td>3.069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physics</td>
<td></td>
<td>14</td>
<td>1.103</td>
<td>0.847 - 1.358</td>
<td>25.251</td>
</tr>
<tr>
<td>Chemistry</td>
<td></td>
<td>15</td>
<td>1.342</td>
<td>1.093 - 1.591</td>
<td>58.709</td>
</tr>
<tr>
<td>Biology</td>
<td></td>
<td>13</td>
<td>1.037</td>
<td>0.772 - 1.302</td>
<td>48.434</td>
</tr>
</tbody>
</table>

* $p < .05$,

There were no statistically significant differences among the mean effect sizes of the studies on physic, chemistry and biology subject areas ($Q_{bet} = 3.069; p = .215$). Considerably strong effects were found for chemistry studies ($d = 1.342$).

Table 4. Effect sizes for mode of instruction

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect ($Q_{bet}$)</th>
<th>n</th>
<th>Mean d</th>
<th>$95%$ CI for d</th>
<th>Homogeneity within Each class $Q_{wi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mode of instruction</strong></td>
<td>2.712</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td></td>
<td>26</td>
<td>1.264</td>
<td>1.075 - 1.453</td>
<td>99.919</td>
</tr>
<tr>
<td>Combination</td>
<td></td>
<td>16</td>
<td>1.006</td>
<td>0.764 - 1.248</td>
<td>38.007</td>
</tr>
</tbody>
</table>

* $p < .05$,

There were no statistically significant differences between the mean effect sizes of studies using conceptual change text and mixed studies ($Q_{bet} = 2.712$, $p = .099$). Also it was found that CCT model has been more successful than mixed studies (CCT combined with other instruction) in promoting the students’ academic success.

Table 5. Effect sizes for school level

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect ($Q_{bet}$)</th>
<th>n</th>
<th>Mean d</th>
<th>$95%$ CI for d</th>
<th>Homogeneity within Each class $Q_{wi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School level</strong></td>
<td>2.739</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td>8</td>
<td>1.077</td>
<td>0.743 - 1.411</td>
<td>24.578</td>
</tr>
<tr>
<td>Secondary</td>
<td></td>
<td>25</td>
<td>1.107</td>
<td>0.913 - 1.302</td>
<td>73.510</td>
</tr>
<tr>
<td>University</td>
<td></td>
<td>9</td>
<td>1.399</td>
<td>1.085 - 1.712</td>
<td>34.390</td>
</tr>
</tbody>
</table>

* $p < .05$,

There were no statistically significant differences among the mean effect sizes of the studies conducted in primary school, secondary school and university levels ($Q_{bet} = 2.739$, $p = .254$) and CCT instruction was the most effective in university level.
Table 6. Effect sizes for sample size

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect (Q_{bet})</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d</th>
<th>Homogeneity within Each class Q_{wi}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n ≤ 25</td>
<td>4.170</td>
<td>12</td>
<td>0.975</td>
<td>0.688 - 1.261</td>
<td>15.617</td>
</tr>
<tr>
<td>25 &lt; n ≤ 50</td>
<td></td>
<td>27</td>
<td>1.195</td>
<td>1.012 - 1.378</td>
<td>87.790</td>
</tr>
<tr>
<td>50 &lt; n ≤ 75</td>
<td></td>
<td>3</td>
<td>1.577</td>
<td>1.053 - 2.101</td>
<td>28.447</td>
</tr>
</tbody>
</table>

*p < .05,

Also it was found that there were no statistically significant differences in mean effect sizes of the studies with respect to sample size (Q_{bet}= 4.170, p= .124).

Table 7. Effect sizes for nature of assessment instrument

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect (Q_{bet})</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d</th>
<th>Homogeneity within Each class Q_{wi}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of assessment instrument</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimenter developed</td>
<td>0.379</td>
<td>29</td>
<td>1.168</td>
<td>0.983 - 1.353</td>
<td>86.520</td>
</tr>
<tr>
<td>Standard</td>
<td></td>
<td>3</td>
<td>1.000</td>
<td>0.410 - 1.591</td>
<td>4.105</td>
</tr>
<tr>
<td>Combination</td>
<td></td>
<td>10</td>
<td>1.210</td>
<td>0.899 - 1.521</td>
<td>53.351</td>
</tr>
</tbody>
</table>

*p < .05,

It was found that there were no statistically significant differences in mean effect sizes of the studies with respect to nature of assessment instrument (Q_{bet}= 0.379, p= .826).

Table 8. Effect sizes for type of instrument

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect (Q_{bet})</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d</th>
<th>Homogeneity within Each class Q_{wi}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of instrument</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple choice test</td>
<td>0.111</td>
<td>27</td>
<td>1.163</td>
<td>0.971 - 1.356</td>
<td>84.642</td>
</tr>
<tr>
<td>Two tier test</td>
<td></td>
<td>10</td>
<td>1.142</td>
<td>0.822 - 1.462</td>
<td>50.926</td>
</tr>
<tr>
<td>Three tier test</td>
<td></td>
<td>5</td>
<td>1.233</td>
<td>0.793 - 1.674</td>
<td>9.115</td>
</tr>
</tbody>
</table>

*p < .05,

It was found that there were no statistically significant differences in mean effect sizes of the studies with respect to type of instrument (Q_{bet}= .111, p=.945).

Table 9. Effect sizes for publication date

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect (Q_{bet})</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d</th>
<th>Homogeneity within Each class Q_{wi}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995-1999</td>
<td>1.654</td>
<td>3</td>
<td>0.978</td>
<td>0.421 - 1.536</td>
<td>2.735</td>
</tr>
<tr>
<td>2000-2005</td>
<td></td>
<td>14</td>
<td>1.075</td>
<td>0.813 - 1.337</td>
<td>47.637</td>
</tr>
</tbody>
</table>
It was found that there were no statistically significant differences in mean effect sizes of the studies with respect to the publication date ($Q_{bet}= 1.654, p= .43733$).

**Table 10.** Effect sizes for duration of treatments

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect ($Q_{bet}$)</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d Lower</th>
<th>%95 CI for d Upper</th>
<th>Homogeneity within Each class $Q_{ws}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Duration of treatments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4 weeks</td>
<td></td>
<td>2</td>
<td>1.202</td>
<td>1.020</td>
<td>1.388</td>
<td>98.618</td>
</tr>
<tr>
<td>4-6 weeks</td>
<td></td>
<td>27</td>
<td>1.014</td>
<td>0.722</td>
<td>1.306</td>
<td>24.422</td>
</tr>
<tr>
<td>6-8 weeks</td>
<td></td>
<td>11</td>
<td>1.442</td>
<td>0.946</td>
<td>1.937</td>
<td>12.654</td>
</tr>
<tr>
<td>Unspecified</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It was found that there were no statistically significant differences in mean effect sizes of the studies with respect to duration of treatments ($Q_{bet}=2.362, p= .306$).

**Table 11.** Effect sizes by instructor effect

<table>
<thead>
<tr>
<th>Variable and class effect</th>
<th>Between class effect ($Q_{bet}$)</th>
<th>n</th>
<th>Mean d</th>
<th>%95 CI for d Lower</th>
<th>%95 CI for d Upper</th>
<th>Homogeneity within Each class $Q_{ws}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instructor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Researcher</td>
<td></td>
<td>14</td>
<td>1.380</td>
<td>1.129</td>
<td>1.632</td>
<td>56.886</td>
</tr>
<tr>
<td>Classroom Teacher</td>
<td></td>
<td>28</td>
<td>1.061</td>
<td>0.884</td>
<td>1.237</td>
<td>72.901</td>
</tr>
</tbody>
</table>

Instructor effect was a factor that affecting students’ academic success. It was found that there were statistically significant differences in mean effect sizes of the studies with respect to instructor effect ($Q_{bet}= 4.159, p= .041$). It was found that researchers were more effective than classroom teachers on performing the method successfully.

**Discussion and Conclusion**

In this meta-analysis, 42 experimental groups having the necessary data to estimate the effect size were determined. Effect sizes were calculated from the 42 studies. The overall effect size for CCTs was calculated as 1.18 of the 95% confidence interval (1.019-1.314). This is a large effect size according to Cohen’s criteria. Moreover, a Z-score of 1.20 corresponds to the 88th percentile in a standardized normal distribution. This means that, the students instructed with CCTs achieved 88% more success than the traditional students. This means that CCTs have been more effective than the traditional instruction. This finding is consistent with the results of some other studies (Al khawaldeh, 2007; Al khawaldeh & Al Olaimat, 2010; Alparslan, Tekkaya & Geban 2003; Çetin, Ertepınar & Geban, 2015; Özkam, Tekkaya & Geban, 2004; Yürük, 2007; Dilber & Düzgün, 2007; Taştan, Yalçınkaya & Boz, 2008b).

It is expected that the average effect of published studies will be greater because the probability of publication of studies giving statistically significant results is higher (Rosenthal, 1991; Cooper and
Hedges, 1994). In this meta-analysis study, 27 published studies (articles) and 15 unpublished studies (master thesis, doctoral dissertations and conference papers) were analyzed. Both published and unpublished studies gave similar results in terms of effect sizes. These findings indicated that there was no publication bias in respect to publication status. It can be indicated with confidence that the sample of studies used in this meta-analysis was representative of the population of studies relating to CCTs.

In this meta-analysis study, 27 articles, 9 master theses, 4 PhD dissertations, and 2 conference papers providing the criteria were analyzed. Effect sizes of all studies were high. When the analyses were examined, it was found that there is no significant difference between the effect sizes of these four groups. It was seen that the conference papers (only two) have the highest effect size. However, more conference papers should be included in order to generalize the results.

In the study, the largest mean effect was found in the chemistry studies (d = 1.342), suggesting that CCTs is most effective in chemistry. Moreover in the fields of physics and biology, a high mean effect was also found, the values were d = 1.103 and d = 1.037, respectively. On the other hand, there were no statistically significant differences among the mean effect sizes of physics, chemistry and biology subject areas (\( \text{Q}_{\text{het}} = 3.069; p = .215 \)). This means that, the effect size of the CCTs does not differ according to the subject matter applied.

In the literature, there are two main trends related to CCTs. First, the studies compare the CCT with the traditional method. Second, the studies compare the CCT accompanied another method with the traditional method. The number of studies which compare the conceptual change approaches among themselves is scarce (Balç, 2005; Okur, 2009; Durmuş and Bayraktar, 2010). Hence, more research is needed to make such comparisons. Thus, it can be determined which technique will yield more effective results for conceptual change. In literature, the studies using two methods (CCTs accompanied by any other method) support the idea that these methods can eliminate each other’s deficiencies and more effective results can be achieved (Guzzetti et al., 1997; Hynd et al., 1997; Harrison and Treagust, 2001; Ünal, 2007; Sevim, 2007; Okur, 2009). For a better understanding of which instruction was the most effective, 26 studies which use CCTs and 16 studies which use both CCTs and additional another technics were analyzed in this study. Considering the effect sizes between these approaches, it was found that there is no significant difference between the effect sizes of these two approaches (\( \text{QB} = 2.712; p = .099 \)). That is to say, the effects sizes of studies do not differ according to the type of instruction used.

To investigate the effect size of the studies with respect to learning level, the studies were divided into three categories; primary, secondary and university. The CCT method was found to be most effective at university level (d = 1.399). This method was also found to be effective at primary and secondary levels as well (respectively d = 1.077 ve d = 1.107). This meta analysis indicated that there is no significant effect size differences between the studies in terms of different school levels. (\( \text{QB} = 2.739; p = .254 \)). Although, it has been found that the university level studies have more effective results, CCT method have an effect for all levels, therefore; this method can also be used in the primary and secondary levels.

It was found that there was no significant difference between the mean effect sizes of the groups according to sample size in the study (\( \text{QB} = 4.170; p = .124 \)). High effect sizes were determined for all sample sizes and these values were calculated as d = 0.975, d = 1.195 and d = 1.577 for low, medium and large learning levels, respectively.

In this meta-analysis study, 29 tests developed by researchers, three standard test, and 10 mix tests were analyzed. When the analyses were examined, it was found that there is no significant difference
between the effects sizes of these three groups. Also, it was found that the tests developed by the researchers are more effective than the standard tests. This result is parallel with another meta-analysis studies examining the effectiveness of the researcher made or standard tests (Bayraktar, 2000; Başol Göçmen, 2003).

The test type was considered as another feature that is considered to have an effect on the meta-analysis work related to the CCTs due to its importance in determining the alternative concepts. Multiple choice test was used at 69% in the studies included in this meta-analysis study. In the literature, it is stated that the use of two-tier or three-tier tests is more effective in determining the misconceptions of students (Eryılmaz and Sürmeli, 2002; Mutlu and Özel, 2008). It was found that, although there was no significant difference between the mean effect sizes of these groups in this study, the studies using the three-tier test had greater effect size (QB = 0.111; p = .945). When the literature related to CCT was examined, it was found that studies were usually used multiple-choice tests. However, it was noticed that two tier or three tier tests have also been used. Two-tier tests consist of two parts. The first stage consists of a multiple choice question and the second stage consists of the part where the explanation of the reason for the selection of the answer to the first stage is desired. The second stage can be a multiple-choice or open-ended question. The three-tier tests consist of three parts. The difference of the three-tier test from the two-tier test is that it is made up of an additional third part that is questioned about the degree of assurance of the answers given to the first two steps. Multiple choice tests are preferred because of their ease of implementation in large samples and their easy analysis of results (Eryılmaz ve Sürmeli, 2002). Though there are advantages of multiple-choice tests, two-tier and three-tier tests are more useful in emerging alternative concepts in students since students explain the reasons of their answers (Dikmenli, Türkmen, Çardak ve Kurt, 2005; Bilgin, 2006; Efe, 2007; Gürdal, 2008; Özdemir, 2008).

When the studies were classified according to the publication year, it was observed that there has been a big increase in the number of studies related to CCTs in the last five years. The effect size of the publications of the last five years had the largest value compared to the other years, whereas the high effect size was determined in all categories of the publishing year. There was no significant difference between the mean effect sizes of the groups (QB = 1.654; p = .437). The fact that the effect size value of recent years was higher indicates that the researches carried out today could planned in detail and made under better conditions. In addition, new studies are designed in line with the proposals of the previous studies and aim to overcome the limitations of older studies.

A large majority (64%) of the studies lasted for 4-6 weeks. High effect size was determined in all categories of the treatment period. However, it was found that there is no significant difference between the mean effect sizes of the durations of the studies (QB = 2.362; p = .306). The effect size was found to be higher for 4-6 weeks when the application period was minimum. This may be a sign that as the duration of application increases, the students become bored and the number of students decreased, and therefore the effect of the method may be diminished.

In this study, the influence of the practitioners was determined as the only significant factor by affecting student achievement. According to the analysis results, it was determined that there were statistically significant differences between the mean effect sizes of the groups. It was found that the researcher practitioner is more effective than the classroom teacher practitioner. This suggested that the researcher is more prevalent in the subject or that sufficient information or training on the practice may not be given to the classroom teacher.
Suggestions

The suggestions regarding the results of meta-analysis:
1. The detailed information about researchers should be given in the studies regarding CCTs.
2. The implementer and the researcher should not be the same person in the studies regarding CCTs.
3. The implementer should be educated about the implementation by the researcher.
4. The duration of implementation should clearly be stated.
5. The professional experience of the implementer and his/her pedagogical content knowledge should be expressed clearly.
6. The studies about CCTs should be conducted on primary and university students.
7. The studies about CCTs accompanied by analogy, concept map, computer assisted instruction should be conducted.
8. Retention tests should be administered in the experimental studies after the administration of pre- and post-tests in order to increase the reliability of study.
9. The implementation of CCTs in the classroom should be given enough time.
10. The ideas of students should be taken into account on how to implement CCTs in order to increase the effect of CCTs.

Suggestions regarding the researchers
1. The future researchers should evaluate the effect size of their studies comparing the effect size of this study (d=1.18).
2. The researchers who will conduct a meta-analysis study should contact the researcher who did not give detailed information about his or her study.
3. Unpublished researchers should be included to the meta-analysis studies.
4. The meta-analysis programs (e.g. Metawin 1.0) should be compared in terms of advantages and disadvantages of them.
5. More meta-analysis research should be conducted to examine the effect of CCTs on attitude.
6. More meta-analysis research should be conducted to examine the effect of analogy, computer assisted instruction.
7. This study states that the number of studies which use two or three tier tests is quite few. Therefore, these tests should be used in the studies about CCTs and the studies examining the students’ misconceptions.

The suggestions regarding science education and curriculum development
1. The results of meta-analysis studies should be used in the studies of curriculum development.
2. The curriculum developers should suggest the usage of CCTs to the chemistry, physics, and biology teachers since these studies have large effect size.
3. In-service training courses about CCTs should be arranged for the science teachers. Also, a course about CCTs should be defined for the education faculties.
4. Textbook writers and textbook evaluation committee should pay attention to the concept teaching in order to eliminate misconceptions of students.
5. The studies planned to be done on CCTs should be considered as priority by Ministry of National Education and The Scientific and Technological Research Council of Turkey.

References


Bibliography of Studies Included in Meta-Analysis


Ruhur, Ö., & Geban, Ö. (2002). Effectiveness of conceptual change text oriented instruction on students' understanding of cellular respiration concepts. Biochemistry and Molecular Biology Education. 30(4), 239–243


