A Meta-Analysis of the Self-Explanation Effect

Rebecca Mugforda, Shevna Coreya, Craig Bennella, & Catherine Martensb

Department of Psychology, Carleton Universitya; Criminology Department, University of Ottawab

Ottawa, Ontario, Canada

Abstract

The current study represents the first meta-analytic review of the self-explanation effect, the instructional practice of having students explain learning materials to themselves in order to enhance learning outcomes. Meta-analytic techniques were applied to 42 independent effects. Results indicate a moderate effect of self-explanation on near transfer test scores ($g = .39$). Although significant heterogeneity of effects was found, moderators selected for inclusion in the current meta-analysis failed to account for this variability. Potential reasons for these findings as well as future research directions are discussed.

Introduction

Self-explanation (SE) is a germane load-inducing activity that refers to the process of generating statements during learning in order to actively make sense of instructional materials (Chi, 2000). The SE effect (that learning is enhanced through SE) has been investigated in a wide variety of educational domains, including mathematics (Akersten et al., 2003), biology (Haussmann & Chi, 2002), and physics (Conati & VanLehn, 1999).

Researchers have postulated that the beneficial effects of SE stem from the fact that it promotes a more deliberate processing of instructional materials, thus enhancing schema acquisition (Chi et al., 1994).

Although research has generally supported the use of SE as an instructional aid (Akersten et al., 2003; de Bruin et al., 2007; Renkl et al., 1998), some studies have produced less favourable results (Grobe & Renkl, 2006; Gross & Renkl, 2007; Mwandi & Sweller, 1998).

The overall magnitude of the SE effect is currently unknown. Furthermore, little is known about potential factors that may influence the SE effect.

Hypotheses

This meta-analysis attempts to provide answers to two questions: (1) what is the overall effect of SE, and (2) are there identifiable differences between studies that may contribute to the variability in the SE effect?

Based on a review of the literature, the following two hypotheses were of central concern:

**Hypothesis 1.** Those who use SE will have higher test scores than those who do not.

**Hypothesis 2.** The strength of the SE effect will be moderated by the type of SE. Specifically, prompted SE will be superior to spontaneous or instructional SE. Further, it is expected that spontaneous SE will be superior to instructional SE.

Other potential moderators include: instructional format, type of population, level of element interactivity, field of study, pace of learning, and whether or not feedback is received.

Method

**Sample**

Studies were found by searching databases, reference sections of published studies, and by contacting key researchers to obtain unpublished manuscripts or data.

The following inclusion criteria were used:

1. The study must have an experimental group engaging in some form of SE and a control group not engaging in SE.
2. The study must have a dependent measure related to learning outcomes.
3. The study must have enough information to allow for the calculation of an effect size.

Studies were coded on the basis of each of the potential moderators listed above.

Results

**Overall**

A moderate effect of self-explanation on near transfer test scores was found ($g = .39$, 95% CI = .26 to .51).

The homogeneity statistic was also significant, $Q(41) = 70.03, p < .01$, indicating that one or more variables moderate this effect. As shown in Table 1, however, the moderators selected for inclusion in the current meta-analysis did not account for this variation ($p > .05$).

<table>
<thead>
<tr>
<th>Moderator</th>
<th>$g$ (SE)</th>
<th>$CI_95$</th>
<th>$Q(df, d)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of SE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prompted</td>
<td>.39 (.08)</td>
<td>.24 to .54</td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>.50 (.18)</td>
<td>.15 to .85</td>
<td></td>
</tr>
<tr>
<td>Instructional</td>
<td>.24 (.17)</td>
<td>.11 to .40</td>
<td></td>
</tr>
<tr>
<td>Type of Population</td>
<td></td>
<td></td>
<td>1.80 (2)</td>
</tr>
<tr>
<td>Post-Secondary</td>
<td>.38 (.08)</td>
<td>.22 to .54</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>.54 (.15)</td>
<td>.24 to .84</td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>.26 (.15)</td>
<td>.03 to .55</td>
<td></td>
</tr>
<tr>
<td>Element Interactivity</td>
<td></td>
<td></td>
<td>0.04 (1)</td>
</tr>
<tr>
<td>High</td>
<td>.38 (.07)</td>
<td>.25 to .51</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.43 (.24)</td>
<td>.04 to .91</td>
<td></td>
</tr>
<tr>
<td>Field of Study</td>
<td></td>
<td></td>
<td>0.31 (3)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>.36 (.09)</td>
<td>.18 to .54</td>
<td></td>
</tr>
<tr>
<td>Engineering/technical</td>
<td>.40 (.17)</td>
<td>.09 to .78</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>.45 (.10)</td>
<td>.16 to .73</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>.37 (.19)</td>
<td>.00 to .74</td>
<td></td>
</tr>
<tr>
<td>Pacing of Learning</td>
<td></td>
<td></td>
<td>0.00 (1)</td>
</tr>
<tr>
<td>Limited</td>
<td>.51 (.13)</td>
<td>.26 to .76</td>
<td></td>
</tr>
<tr>
<td>Self Paced</td>
<td>.50 (.08)</td>
<td>.35 to .65</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
<td></td>
<td>0.03 (1)</td>
</tr>
<tr>
<td>Yes</td>
<td>.37 (.16)</td>
<td>.09 to .64</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>.39 (.07)</td>
<td>.24 to .54</td>
<td></td>
</tr>
</tbody>
</table>

Note: $N =$ number of effect sizes, $g =$ Hedge's effect size; $SE =$ standard error; $CI_95 =$ 95% confidence interval around the effect size; $CI =$ variability between the categories of moderators; $d =$ degrees of freedom.

Discussion

This study represented an initial attempt to estimate the overall effect of SE on learning and to determine what factors contribute to the variability in the SE effect.

**Hypothesis 1**

Overall, the hypothesis that SE is superior to no SE was supported. Indeed, the current meta-analysis found a moderate positive effect of SE on near transfer test scores across a variety of educational domains and instructional formats.

**Hypothesis 2**

The moderating effect of type of SE, however, was not found in the current meta-analysis.

While the homogeneity statistics was significant, none of the other potential moderators that were examined were identified for variations in the SE effect either. Presumably, other moderators, which were not examined in the current study, explain the variation in the SE effect across studies (see below).

Limitations and Future Directions

(1) Large amount of within-study variability

Across all moderators, the Co associated with the estimated effects were relatively large. Several factors may be contributing to this, including large standard errors (within study error) and relatively small sample sizes (resulting in insufficient power to detect significant moderating effects).

Similarly, although the expected pattern of results with respect to type of SE was partially found (i.e., effect sizes associated with instructional SE were smaller than studies involving prompted or spontaneous SE, the larger effect for spontaneous SE was not expected. Given the fact that the average effect of spontaneous SE was based on a small number of effects ($k = 6$), and that a large amount of variation existed within these studies themselves (wide CI), results of this moderator analysis should be interpreted with caution.

Further inspection of the SE literature reveals that if a meta-analysis was conducted on studies directly comparing the effects of prompted SE to spontaneous SE, the expected pattern of results would likely emerge.

(2) The presence of more relevant moderator variables

As indicated above, a plausible explanation for the lack of significant moderators is that other, more relevant moderators may exist. For instance, the experience level of participants may moderate the effect of SE (e.g., Renkl et al., 1998). This was not investigated in the current meta-analysis, however, as few studies explicitly stated the experience level of participants.

Another potential moderator may be the type of SE instructions provided to participants (i.e., type of prompt). There is evidence that the type of prompt can differentially affect learning outcomes (e.g., Berthold et al., in press; Schwerin & Renkl, 2007). Indeed, certain types of prompts (e.g., detailed, principle-based prompts; Hilbert & Renkl, in press) may elicit higher quality SE than other types of prompts (e.g., less detailed, open prompts; Berthold et al., in press), which may ultimately lead to greater learning. This was not investigated in the current meta-analysis because SE prompts varied widely between studies, making it difficult to create meaningful moderator categories.

Future meta-analyses are planned to disentangle these sorts of issues.

References


