Socrates in the schools: Gains at three-year follow-up

Frank Fair[a]

psy_fkf@shsu.edu

Lory E Haas[a]

leh015@shsu.edu

Carol Gardosik[b]

gardosikc@stthom.edu

Daphne Johnson[a]

ddjohnson@shsu.edu

Debra Price[a]

EDU_DPP@shsu.edu

Olena Leipnik[a]

SOC_OVL@shsu.edu

Abstract

Three recent research reports by Topping and Trickey (2007a; 2007b), by Fair and colleagues (2015), and by Gorard, Siddiqui and Huat See (2015) have produced data that support the conclusion that a Philosophy for Children (P4C) program of one-hour-per-week structured discussions has a marked positive impact on students. This article presents data from a follow up study done three years after the completion of the study reported in Fair et al. (2015). The data show that the positive gains in scores on the Cognitive Abilities Test (CogAT in the USA and CAT in the UK) were still present and had not faded after three years. Given the strength of these confirmations of the positive durable impact of the P4C program of structured discussions and given the relatively low cost of implementing the P4C program, it is recommended that it become a standard part of the school curriculum.

Key Words

cognitive abilities, community of inquiry, Philosophy for Children

[a] Sam Houston State University, Texas
[b] University of St Thomas, Texas

Introduction

After successfully replicating the results of an earlier Scottish study of the effect of a P4C program of one-hour-per-week philosophy discussions on students' scores on the Cognitive Abilities Test (CogAT in the USA, CAT in the UK), our research group in Texas extended the replication by doing a three-year follow-up study to determine if the relative improvements in the CogAT/CAT scores for the experimental group over the control group that were visible in the original 2010-2011 study were still detectable in the spring of 2014.

The original studies done in Scotland by Topping and Trickey

In 2009, when we formed our Philosophy for Children (P4C) interest group, we looked for hard evidence for the educational impact of P4C programs. When we found Topping and Trickey's study of the effects of a P4C program on students in Clackmannanshire, Scotland (Topping & Trickey 2007a), our Philosophy for Children (P4C) study group decided to replicate their study in a middle school in Texas. We had several reasons for our choice. First, Topping and Trickey used the Cognitive Abilities Test (CogAT in the US, CAT in the UK) as their tool to measure the impact of the P4C program, and thus they used an instrument that was well-regarded and widely-studied. From the publisher's website here is a description of the CogAT/CAT:

> Reasoning abilities have substantial correlations with learning and problem solving, both in and out of school. *CogAT's* measurement of three different content domains ensures that educators receive a balanced view of the child.
>
> Each level of *CogAT* offers three test batteries:
>
> - Verbal
> - Quantitative
> - Nonverbal (Riverside Publishing 2011)

Scores on the CogAT/CAT have been the subject of some study, most notably of how those scores correlate well with scores on General Certificate of Secondary Education (GCSE) exams given in a variety of subjects such as English, Science, Mathematics, and History to 15-16 year-old students in England, Wales, and Northern Ireland (Deary, Strand, Smith & Fernandes 2007). Topping and Trickey mention this correlation in their report (2007a) and, in effect, use the CogAT/CAT scores as proxies to measure a positive impact on academic achievement (the interested reader can find further information about the CogAT/CAT by visiting

http://faculty.education.uiowa.edu/david-lohman/home which contains numerous papers by the test developer Dr David Lohman of the University of Iowa). Second, Topping and Trickey's study employed a control group and took care to approximate the conditions of a randomised controlled clinical trial for the CogAT/CAT pretests and posttests. Third, it was very important that the P4C intervention they employed was *cheap*. It was cheap both in terms of instructional time (one hour per week) and in terms of the materials used (Cleghorn & Baudet 2002, *Thinking through philosophy: Book 4*, at about $25.00 USD per copy). Nor did it take extensive and expensive staff preparation. All that was done was to take the participating teachers, those who had been randomly selected, through one day on training with the materials. The training involved an introductory overview of P4C, and then working through three lessons in the book in the same was that they would in turn work with their students.

The point is simply that an educational intervention that is costly in terms of instructional time, materials, or teacher training is less likely to be adopted as a standard component of the school curriculum even if it has a beneficial effect on students. We were looking for a program that school leaders would be more likely to adopt once its educational effectiveness was demonstrated, and, with that aim in mind, less costly is better.

In the Scottish study, Dr Keith Topping of the University of Dundee supervised the work of a doctoral student, Steven Trickey, who was at that time working as a psychologist for the Clackmannanshire Educational Authority where the study took place. In addition, Paul Cleghorn served as a Head Teacher in the school system and, along with Stephanie Baudet, had put together a set of materials for reflection in a series of books entitled *Thinking through philosophy*. This particular study used *Thinking through philosophy: Book 4* and that book contained an extensive description of how to engage students to form a community of inquiry in the classroom. Each philosophy discussion session was structured as a process with a series of seven stages, beginning with a focusing exercise to create a relaxed, meditative frame of mind and followed by a brief link with the previous week's discussion. Then came the stimulus for discussion—usually reading a story from *Thinking through philosophy*. Next the students worked in pairs discussing the story and reflecting on some open-ended questions suggested by the story. This was followed by dialogue in larger groups where the teacher has encouraged the students to form a *community of inquiry* by:

(a) communicating their views in response to the questions at hand,

(b) supporting their views with reasons,

(c) listening respectfully to the views of others,

(d) indicating whether they agree or disagree with those views,

(e) providing alternative viewpoints, and

(f) gradually developing a process of dialogue.

Finally, the teacher brings closure by encouraging the students to reflect on the discussion and how their thinking might have progressed, and by providing a 'thought for the week' that highlights an idea to serve as a basis for 'homework' to be reflected on in order to relate the idea to situations outside the original stimulus (Topping & Trickey 2007b; for more detail see Cleghorn & Baudet 2002).

Topping and Trickey (2007a) reported achieving substantial pretest to posttest gains in CogAT/CAT scores for their experimental group versus their control group, and our P4C interest group decided those results were worth serious attention. But then Topping and Trickey did what few others had done, namely, they conducted a follow-up study two years later in 2005. This was important because of the tendency of the impact of many educational interventions to fade over time (Cascio & Staiger 2012). In their follow-up study Topping and Trickey (2007b) were clearly able to document the persistence and durability of the effects on the students' cognitive abilities of the P4C intervention, and that was very encouraging, given the correlation of CogAT/CAT scores with academic achievement as documented by Deary et al. (2007) and others. The results of the original study and the two-year follow-up can be seen in Figure 1 provided to us by Dr Trickey.

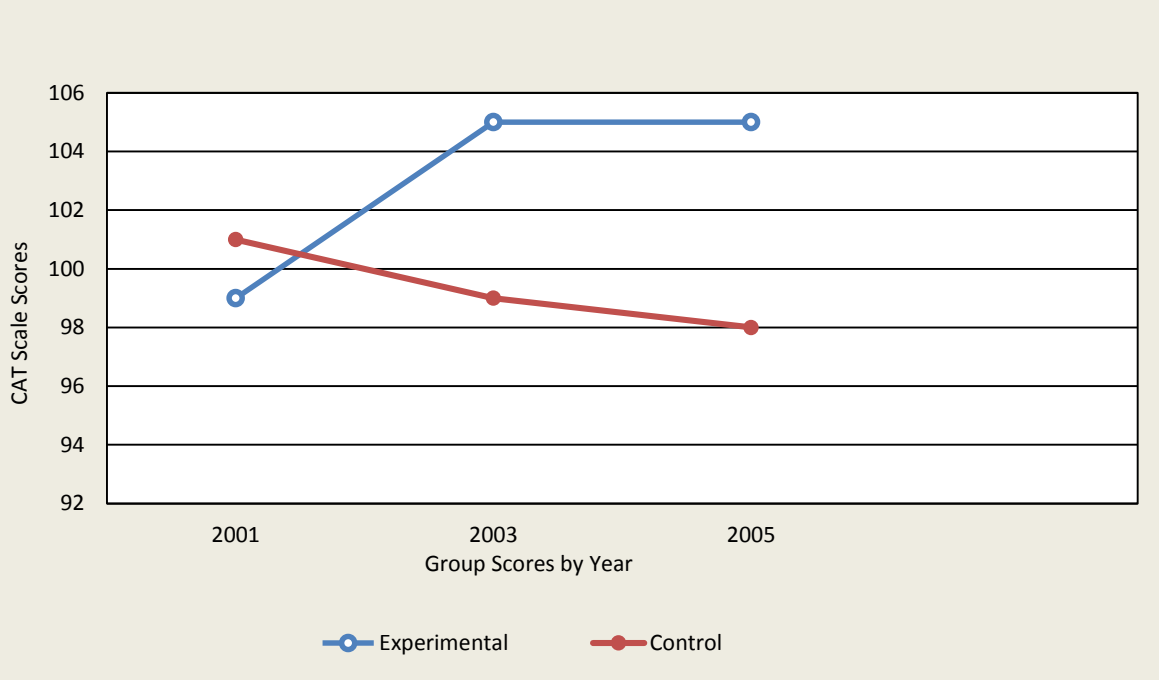

Figure 1: Students' scores on the Cognitive Abilities Test from the Scottish study

I. Our initial study in 2010-2011

In this section we provide a brief description of our 2010-2011 initial study in order to provide background information for our report of our follow-up study three years later. If a more complete description of the initial study is desired, please see the article which reports on it (Fair et al. 2015). During the school year of 2010-2011, we attempted to replicate the Topping and Trickey (2007a) study by looking at the impact of the same P4C program on seventh graders (12-13 years old) in a Texas middle school. We structured the situation to emulate, as closely as we could, the intervention that Topping and Trickey employed in their study. Thus we used the same materials from *Thinking through philosophy: Book 4* as the main source to prompt the weekly discussions. In addition, we spent one day before the start of school with the teachers to train them by having them role-play as students. We followed the same 7-stage process for each discussion. Prior to the start of the school term in late August of 2010, the teachers were trained by having them role-play as students. Dr Trickey was able to consult with us and take part in the teacher-training session. The participating seventh grade teachers were Language Arts teachers who were randomly selected on the morning of the training. Those seventh grade teachers who were not selected were strongly advised not to incorporate any of the P4C program into their classes in order to maintain the integrity of the control group. We structured the student sessions to emulate, as closely as we could, the intervention

that Topping and Trickey employed in their study. We used the same materials from *Thinking through philosophy: Book 4* as the main source to prompt the weekly discussions. We followed the same 7-stage process for each philosophical discussion. The result was an experimental group containing 186 seventh grade students and a control group containing 79 seventh grade students (for further details, see the report by Fair et al. 2015).

Analyses of the quantitative data

To determine the extent of differences in pretest and posttest scores on the CogAT within the two groups, parametric dependent samples *t*-tests were conducted for both groups of seventh grade participants. For participants in the control group the paired samples analysis yielded a statistically significant difference in pretest and posttest scores, $t(78) = -2.56$, $p = .01$, Cohen's $d = 0.22$, a small effect size. For participants in the experimental group a statistically significant difference in pretest and posttest scores was indicated, $t(185) = -15.40$, $p < .001$, Cohen's $d = 0.65$. According to Cohen (1988), the effect size for the difference was moderate to large. The results of our replication study are displayed in Figure 2.

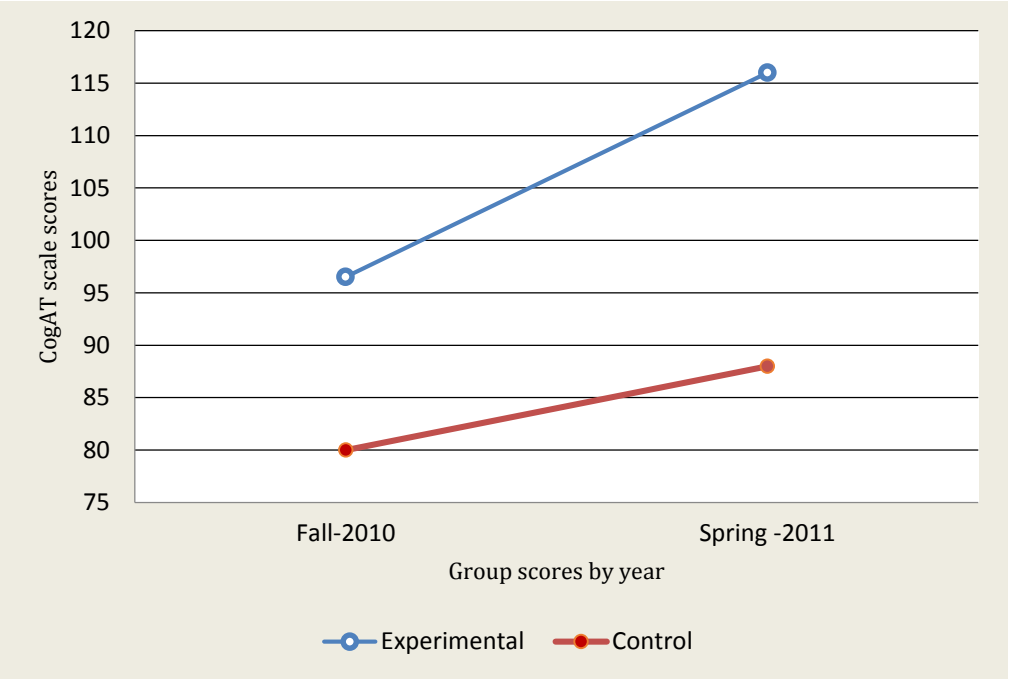

Figure 2: CogAT scores for seventh grade students after 22-26 weeks of P4C

As one can see—and the statistical analysis bears this out—the 22-26 weeks of one-hour-per-week Philosophy discussions within the seventh grade experimental group had a significant impact on CogAT/CAT scores (Fair et al. 2015). This meant that our results for the seventh graders replicated the results that Topping and Trickey (2007a) had obtained. Additionally, our results showed that 22-26 weeks was a sufficient period of time to produce the results, and that length of time could be easily accommodated in one school year in the USA. In contrast, the original Topping and Trickey study concerned the effects of a 58-week program of weekly discussions—somewhat over twice as many weeks as in our study. That led us to wonder whether our program, one involving substantially fewer weeks of discussions, would have a similarly durable impact, or whether the results would be more susceptible to fading over time.

II. The three-year follow-up study

It happened that we were able to conduct our own follow-up study. A large number of students who had been seventh graders during the original study in 2010-2011 were sophomores in the tenth grade, usually 15-16 years of age, and they were in their second year at the local high school during the 2013-2014 school year. The local school district administered the CogAT/CAT to all of the tenth grade students in the spring of 2014, and we were allowed to access the CogAT/CAT score data from the school district archives, specifically the test scores for those sophomores who had either been in the experimental group or the control group in the original study. Furthermore, the scoring of the CogAT/CAT was blind, in that those doing the scoring did not know whether they were scoring follow-up tests from students who had been in the experimental group or students who had been in the control group in the original 2010-2011 study. When we looked at the 2014 score data and used the students' identification numbers from the original study in 2010-2011 to sort them into the two groups, we had the tenth grade CogAT/CAT scores for 133 out of the original 186 members of the experimental group and 50 out of the original 79 members of the control group. These numbers compare well with the numbers that Topping and Trickey were able to obtain for their two-year follow-up—71 out of 105 from their experimental group and 44 out of 72 from their control group (Topping & Trickey 2007b).

The question was whether a follow-up at the three-year mark, instead of the two year mark for the Topping and Trickey study, might display a noticeable 'fading' of the results. Here is what we found.

Analyses of the quantitative data

Analyses of data were conducted to determine whether a statistically significant difference was present between the pretest scores as seventh grade students and posttest scores of experimental group and control group participants three years later as tenth grade students. The reader will note that the means are slightly adjusted as plotted on the previous figure for the first and second round of testing for seventh grade students. This is due to changes in the number of participants in the final round of testing for comparison of all three rounds of testing.

Parametric dependent samples *t*-tests were conducted to determine differences between pretest and posttest scores among the two groups. For participants in the control group the paired samples analysis yielded a statistically significant difference in pretest and posttest scores, $t(49) = -2.67$, $p = .01$, Cohen's $d = 0.28$, a small effect size. For participants in the experimental group a statistically significant difference in pretest and posttest scores was indicated, $t(132) = -10.28$, $p < .001$, Cohen's $d = 0.68$. According to Cohen (1988) the effect size for the difference was moderate to large. In reference to pretest and posttest scores for tenth grade students, a greater difference in posttest scores was present for participants in the experimental group than for participants in the control group, which is consistent with the results noted between the two groups in 2010-2011. Presented in Table 1 are the descriptive statistics for all three rounds of pretest and posttest scores and the results of both studies are presented in Figure 3.

Table 1: Descriptive statistics for pretest and posttest scores by group

	Pretest		**Test**		**Posttest**	
Participants	***M***	**SD**	***M***	**SD**	***M***	**SD**
Experimental ($n = 133$)	100.09	30.41	117.25	27.59	122.53	35.25
Control ($n = 50$)	89.60	37.40	92.4	31.64	100.26	39.09

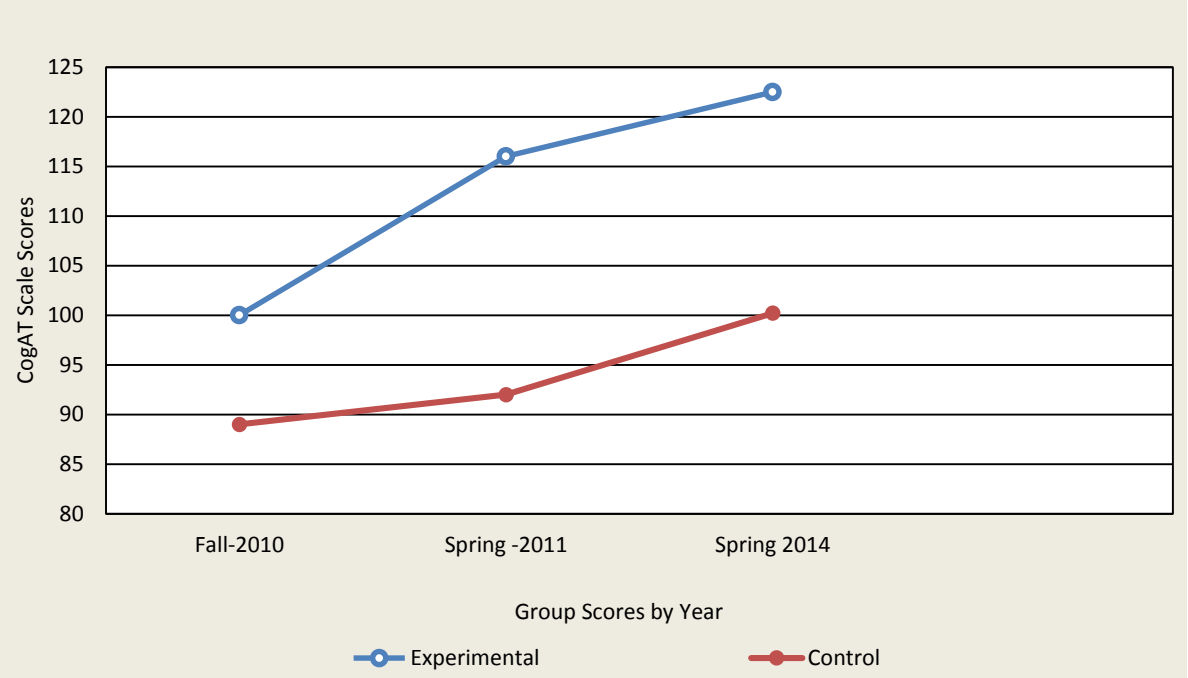

Figure 3: CogAT scores for seventh grade students, and their scores as tenth grade students

III. Discussion

In the article describing the results of their follow-up study, Topping and Trickey write:

> This study provides evidence that gains in cognitive ability which opened up over the intervention period in primary school can be maintained over the subsequent 2 years into secondary school in the same measure, even when pupils have not had any further experience of collaborative inquiry in the secondary school. (2007b, pp. 793-794)

Our study, now including the three-year follow-up report, strengthens and extends their work. First, our study strengthens their results by showing that the program of collaborative philosophical discussions that they used had the same powerful effect on students who were not only half a world away from Scotland, but who were more ethnically diverse. The particular Texas school district enrollment in 2010 was comprised of 46% white, 26.9% African American and 25.8% Hispanic students (Fair et al. 2015). Also the school district is not in a wealthy community since in 2009 the median annual income was US $33,359 in contrast to the state median income of US $49,433. Indeed, at the time of the study 60.2% of the students in the district were classified as 'economically disadvantaged' (Fair et al. 2015). Furthermore, our study

strengthens Topping and Trickey's results by showing that, instead of 58 weeks, fewer weeks (22-26) would suffice to produce the effects. Next, our study demonstrates the *portability* of the program, since, while Dr Trickey did help with the teacher training day, in contrast to the situation in Scotland neither he nor Paul Cleghorn, the author of the discussion materials, was present in Texas during the weeks the program ran. In our replication the philosophy discussions were conducted in local public school seventh grade classrooms by seventh grade Language Arts teachers. Neither the teachers nor the students were hand selected, but nevertheless they succeeded in replicating the results of the Scottish study, showing that the presence of the originators of the program was not essential for the program to have an effect. Finally, as we report now, the fact that we could detect a significant difference between the experimental and control groups of students after *three years*, not just two, gives greater credence to the likelihood that the intellectual effects produced by the one-hour-per-week philosophy discussions are as durable as one could wish.

IV. Conclusion

Historically, the Philosophy for Children movement was inspired by thinkers such as John Dewey and Lev Vygotsky, but the lion's share of the credit in recent years must be given to the pioneering work of Matthew Lipman. Lipman, along with colleagues such as Margaret Sharp and Frederick Oscanyan, established the Institute for the Advancement of Philosophy for Children (IAPC) at Montclair State University (Lipman 2003; Lipman, Sharp & Oscanyan 1980). Since the IAPC was founded in 1974 (for timeline, see IAPC n.d.) other programs have come into existence, both in the USA and abroad, but Philosophy for Children remains, in the American context, a very peripheral enterprise as far as the public schools are concerned. We hope that by confirming and strengthening the findings of Topping and Trickey's seminal study, we can give new energy to the movement to incorporate philosophical discussion in schools in Texas and elsewhere. We also take note here of results reported recently in a study done for the Education Endowment Foundation in the UK. Here are two key conclusions from the report:

1. There is evidence that P4C had a positive impact on Key Stage 2 attainment. Overall, pupils using the approach made approximately two additional months progress in reading and math.

2. Results suggest that P4C had its biggest positive impact on Key Stage 2 results among disadvantaged students (those eligible for free school meals). (Gorard et al. 2015, p. 3).

Given the results of these studies, the powerful impact of the process of one-hour-per-week philosophy discussions is clear, and, since the particular program that structures this process of discussion is not costly, we are led to recommend that the program be incorporated into the school curriculum as a standard component.

References

Cascio, E & Staiger, D (2012) *Knowledge, tests, and fadeout in educational interventions (Working paper 18038)*. National Bureau of Economic Research, Cambridge, MA. Available from http://www.nber.org/papers/w18038 (accessed 15 June 2015).

Cleghorn, P & Baudet, S (2002) *Thinking through philosophy: Book 4*. Educational Printing Services, Blackburn, UK.

Cohen, J (1988) *Statistical power analysis for the behavioral sciences*. 2nd edn, Lawrence Erlbaum, Hillside, NJ.

Deary, I, Strand, S, Smith, P & Fernandes, C (2007) Intelligence and educational achievement. *Intelligence,* 35, pp. 13-21.

Fair, F, Haas, L, Gardosik, C, Johnson, D, Price, D & Leipnik, O (2015) Socrates in the schools from Scotland to Texas: Replicating a study on the effects of a philosophy for children program. *Journal of Philosophy in Schools,* 2(1), pp. 18-37.

Gorard, S, Siddiqui, N & Huat See, B (2015) *Philosophy for Children: Evaluation report and executive summary*. Education Endowment Foundation, Millbank, UK.

IAPC (n.d.) *IAPC timeline.* Montclair State University, Montclair, NJ. Available from http://www.montclair.edu/cehs/academics/centers-and-institutes/iapc/timeline/

Lipman, M (2003) *Thinking in education*. 2nd edn, Cambridge University Press, New York, NY.

Lipman, M, Sharp, AM & Oscanyan, F (1980) *Philosophy in the classroom*. Temple University Press, Philadelphia, PA.

Riverside Publishing (2011) *Cognitive Abilities Test™*. The Riverside Publishing Company. Available from http://www.riversidepublishing.com/products/cogAt/details.html

Topping, K & Trickey, S (2007a) Collaborative philosophical enquiry for school children: Cognitive effects at 10–12 years. *British Journal of Educational Psychology,* 77, pp. 271–288.

Topping, K & Trickey, S (2007b) Collaborative philosophical inquiry for schoolchildren: Cognitive gains at 2-year follow-up. *British Journal of Educational Psychology,* 77, pp. 787-796.