Effects of individual direct-instruction tutoring on foster children's academic skills: A randomized trial

ARTICLE in CHILDREN AND YOUTH SERVICES REVIEW · JUNE 2012
Impact Factor: 1.27 · DOI: 10.1016/j.childyouth.2012.01.036

CITATIONS 10
READS 112

5 AUTHORS, INCLUDING:

Robert J. Flynn
University of Ottawa
38 PUBLICATIONS 239 CITATIONS
SEE PROFILE

Marie-Pierre Paquet
Université du Québec en Outaouais
4 PUBLICATIONS 17 CITATIONS
SEE PROFILE

Tim Aubry
University of Ottawa
97 PUBLICATIONS 605 CITATIONS
SEE PROFILE

Available from: Robert J. Flynn
Retrieved on: 05 October 2015
Effects of individual direct-instruction tutoring on foster children’s academic skills: A randomized trial

Robert J. Flynn a,b,⁎, Robyn A. Marquis a, Marie-Pierre Paquet b,1, Lisa M. Peeke b, Tim D. Aubry a,b

a School of Psychology, University of Ottawa, Canada
b Centre for Research on Educational and Community Services, University of Ottawa, Canada

ARTICLE INFO
Available online 3 February 2012
Keywords:
Tutoring
Direct instruction
Educational achievement
Foster children
Randomized trial
Teach Your Children Well

ABSTRACT
We conducted a randomized effectiveness trial to test the hypothesis that foster children of primary-school age who were exposed to an individualized direct-instruction tutoring intervention delivered by their foster parents would experience significantly greater pre-test to post-test gains in reading and math than would foster children in a wait-list control group. The sample consisted of 77 foster children in 9 local Children’s Aid Societies in Ontario, Canada. At the pre-test, the foster children were aged 6 to 13 years (M = 10.7 years, SD = 1.6) and were in primary-school grades 2 through 7. Forty-two foster children were randomly assigned to the experimental (tutoring) group and 35 to the wait-list control group. The sub-tests of the Wide Range Achievement Test—Fourth edition (WRAT4; Wilkinson & Robertson, 2006) served as the outcome measures. At the post-test, the foster children in the experimental group had made statistically and practically greater gains than those in the control group on the WRAT4 sub-tests of Sentence Comprehension (Hedges’ g = 0.38, p = .05), Reading Composite (g = 0.29, p < .10), and Math Computation (g = 0.46, p < .01) but not on Word Reading (g = 0.19, ns) or Spelling (g = −0.08, ns). The implications of the results for improving foster children’s educational achievement were discussed.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Research in many countries indicates that one of the greatest needs of many young people living in out-of-home care (hereafter, ‘in care’) is improvement in their educational achievement (Flynn, Ghazal, Legault, Vandermeulen, & Petrick, 2004; Jackson, 2007; Trout, Hagaman, Casey, Reid, & Epstein, 2008). Studies have consistently shown that many young people in care experience serious difficulties in primary and secondary school, often lagging one or two years behind their peers in the general population (Trout et al., 2008). As a result, young people in care are less likely to obtain a high school diploma or enroll in and graduate from post-secondary education (PSE), including university, college, and skilled-trade/apprenticeship programs (Pecora et al., 2010). Moreover, recent research in Sweden has demonstrated that care leavers’ educational success in primary school is strongly related to their subsequent psychological adjustment (Berlin, Vinnerljung, & Hjern, 2011). These findings underline the necessity of increasing the amount of intervention research focused on effective ways of helping young people in care improve their short-term and long-term educational outcomes. The main purpose of the present study was to investigate whether individual direct-instruction tutoring by foster parents would improve their foster children’s basic academic skills in reading and math.

The international literature reveals that many young people in care experience significant academic problems, including inconsistent school attendance, below-grade academic performance, and low scores on standardized tests of academic achievement in reading, writing and mathematics (Jackson, 2007; Mitic & Rimer, 2002; Pantin & Flynn, 2006; Shonk & Cicchetti, 2001; Trout et al., 2008). As a result, they are approximately three times as likely as their peers in the general population to be involved in special education (Trout et al., 2008).

The relatively few published Canadian studies on the educational attainment of children and adolescents in care have found results similar to those in the international literature. Flynn and Biro (1998) reported that young people in care in Ontario had higher rates of grade retention and school suspension than their age peers in the general population. In the long-term Ontario Looking After Children (OnLAC) project, Flynn et al. (2004) compared two groups of looked-after young people, aged 5 to 9 and 10 to 15, respectively, with nationally representative samples of the same age from the general Canadian population. Eighty percent of the looked-after older children and 78% of the younger children were rated by their foster parents as performing educationally in the same range as the lowest

* We thank the Canada Education Savings Plan, Human Resources and Skills Development Canada, Government of Canada, for its financial support of this research. We thank also the lead hands, direct service staff, foster parents, and foster children from the nine local Children’s Aid Societies in Ontario who collaborated with us.

⁎ Corresponding author at: Centre for Research on Educational and Community Services, University of Ottawa, 34 Stewart Street, Ottawa, Ontario, Canada K1N 6N5.
E-mail address: rflynn@uottawa.ca (R.J. Flynn).

1 Marie-Pierre Paquet is now affiliated with the Département de psychologie et psychologie, Université du Québec en Outaouais, Gatineau, Québec, Canada.

DOI: 10.1016/j.childyouth.2012.01.036
third of the comparison groups from the general population (who had been rated by their parents on reading, spelling, math, and overall educational performance). In more recent data from the OnLAC project, Miller, Flynn, and Vandermeulen (2008) found that 68% of their sample of young people in care who were 10–15 years of age had changed schools three or more times for reasons unrelated to normal progression through the school system. Also, the proportion who had repeated a grade increased with age, such that 16% of their sample of 5–9 year olds in care had repeated a grade, compared with 27% of their 10–15 year olds and 32% of their 16–20 year olds.

Given the substantial educational difficulties of children and adolescents in care, it is surprising—and grounds for a call to action by educational and child welfare researchers—that there are so few validated interventions to improve the academic status of children in care. We note that Trout et al. (2008) found, in their comprehensive review of 66 years (1940–2006) of research in the USA on the academic status of young people in care, too few validated programs to be able to include a synthesis of such interventions. Forsman and Vinerljung (2012) were able to find, after a comprehensive search of the literature in English, Swedish, Danish, or Norwegian, only 11 interventions that had aimed at improving the school achievement of children in care of primary-school age (6 to 15 years) and that had also used school achievements as outcome measures and a pre-test/post-test design. Forsman and Vinerljung (2012) found that nine of the 11 studies they located had reported some indication of significant improvement, with tutoring the best supported intervention. These positive results for the tutoring of children in care dovetail with the positive findings of a recent meta-analysis of randomized studies of tutoring of children of primary-school age (K–8) in the general population (Ritter, Barnett, Denny, & Albin, 2009). In their meta-analysis of 21 randomized field trials, comprising 28 different study cohorts and published between 1986 and 2004, Ritter et al. (2009) found that tutoring was significantly effective in improving children's progress on reading overall (Hedges' g = 0.30), reading global (g = 0.26), reading letters and words (g = 0.41), reading oral fluency (g = 0.30), and writing (g = 0.45), with positive but statistically non-significant mean effect sizes also found on reading comprehension (g = 0.18) and mathematics (g = 0.27). The type of tutor (i.e., parent, college-age, or community tutors) was unrelated to the effect sizes of the outcomes. Ritter et al. (2009) concluded that volunteer tutoring has a positive effect on young students' reading skills but that little is known about the effectiveness of tutoring on math outcomes. They also concluded that small randomized field trials were useful for extending the evidence base on tutoring.

With tutoring suggesting itself as a promising approach, we decided to test a tutoring intervention based on direct instruction. The latter is a well-organized, structured, and empirically supported teaching methodology designed to enhance the reading, language, and arithmetic skills of educationally disadvantaged children (Ryder, Burton, & Silberg, 2006). In Project Follow Through, direct instruction was identified as the most effective method for teaching educationally disadvantaged children (Bereiter & Kurland, 1981–1982), and, in a more recent meta-analysis, direct instruction emerged as one of the three most effective school-level models for improving educational outcomes in low-performance urban schools (Borman, Hewes, Overman, & Brown, 2003). Finally, in combination with contingency management, direct instruction has been found effective in improving academic outcomes for young people at risk of school failure (Dolezal, Weber, Evavold, Wylie, & McLaughlin, 2007).

We chose to evaluate Maloney's (1998) Teach Your Children Well (TYCW) tutoring program, for several reasons. TYCW is based on direct instruction principles, was developed in Ontario, and includes a full array of curricular materials for learners at different levels. Together, the present investigation and Harper and Schmidt's study (2012) constitute the first controlled assessments of the impact of Maloney's (1998) TYCW program.

Our research hypothesis was the following: foster children in primary school who are exposed to an individual version of the direct-instruction tutoring program known as Teach Your Children Well and delivered by their foster parents will experience statistically and practically greater pre-test to post-test gains in the basic academic skills of reading and math than will foster children in a wait-list control group. We tested this hypothesis by means of a randomized effectiveness trial that was conducted under real-world conditions, rather than by an efficacy trial that would have required near-ideal, laboratory-like conditions. The study had obtained the prior approval of the Social Sciences and Humanities Research Ethics Board of the University of Ottawa.

2. Method

2.1. Participants

Nine local child welfare organizations, known in Ontario as Children's Aid Societies (CASs) and located in widely dispersed regions of the province, chose to participate in the study. In each CAS, child welfare workers and their supervisors nominated foster child—foster parent pairs as candidates for the study because both the children and parents were seen as likely to agree to participate in the tutoring. Also, the foster parents were viewed as likely to be able to deliver the intervention, and the children were perceived as likely to benefit.

2.1.1. Foster children participants

Seventy-seven children, residing in family foster care, assented to participate and were enrolled in the study. At the pre-intervention assessment (i.e., in September–October, 2008), the 77 foster children were aged 6–13 years (M = 10.7 years, SD = 1.6) and in primary-school grades 2–7 (M = 5.3 years, SD = 1.5). Forty-two were randomly assigned to the tutoring (intervention) group (50.0% male, 50.0% female) and 35 to the wait-list control group (42.9% male, 57.1% female).

In order to participate, each child had to meet several eligibility criteria. Besides being aged 6–13 years and enrolled in grades 2–7, they had to be nominated by their child welfare worker and supervisor as likely to benefit from tutoring by one of their foster parents, fluent in English (the tutoring work books and other materials existed only in English), resident in a foster home or kinship care home (whether run by the local CAS or a private provider), living in a placement assessed as stable by the child welfare worker and supervisor, possessing the legal status of a Crown Ward or Society Ward (and thus likely to remain in care for the duration of the study), and willing to sign an informed assent form. Foster children were excluded from the study if they were living in a group home or if, in the judgment of the child welfare worker and supervisor, they were either very strong students (and thus not likely to need tutoring) or extremely weak students or very behaviourally disturbed (and thus not likely to benefit).

2.1.2. Foster parent participants

Sixty-eight foster parents (63 female, five male) met the study inclusion criteria and were enrolled in the study. Thirty-six were randomly assigned to the experimental group and 32 to the wait-list control group. Of the 42 children in the experimental group, 92.9% were tutored by a female foster parent; of the 34 children in the wait-list control group whose foster parents had indicated their gender on their foster–parent questionnaires, 91.4% were tutored by a female foster parent. The foster parents ranged in age from their 20s to their 60s and over, with most in their 40s and 50s. The majority had completed high school or community college, with only a few having obtained a university degree.

The foster parents had to satisfy the following inclusion criteria: nominated by staff of the local CAS as sufficiently motivated and literate to function as a tutor; able to read well (as assessed by a reading test administered at the pre-test by the research staff); willing to undergo a day of tutoring training and to have their subsequent tutoring training.
performance monitored; committed to tutoring their target foster child (or children) for 3 h per week for 30 weeks during the school year; having access to a computer and the Internet in their foster home (to run the computer-based math software); prepared to communicate regularly with the project staff by e-mail or telephone; and willing to sign an informed consent form.

The foster parents were strongly encouraged to tutor only one foster child, out of the maximum of four permitted in foster homes in Ontario. To maximize recruitment, however, we allowed individual foster parents to offer tutoring to a maximum of two foster children, with the understanding that all foster children would be tutored on an individual basis, both to maximize the potential impact of the intervention and to minimize any potential statistical dependencies in the children's research data that might result from tutoring two children together. Of the foster children in the experimental group, 69% were the only ones tutored in their respective foster homes, versus 71% in the control group.

### 2.2. Power analysis, research design, and random assignment

#### 2.2.1. Power analysis

An a priori power analysis by means of the G*Power 3.1 computer program (Faul, Erdfelder, Lang, & Buchner, 2007) suggested that for an analysis of covariance (ANCOVA) carried out via multiple regression, with a medium effect size \( f^2 = .15 \), a desired power of .80, and an alpha level of .05, we would need a total sample size of 55 foster children. Cohen's (1977) power tables indicated a similar sample size (53 foster children).

#### 2.2.2. Research design and random assignment

We used a pre-test/post-test control-group design. Random assignment to conditions, based on a table of random numbers, took place at the pre-test, immediately after the foster parents and foster children had signed their respective consent or assent forms. We used a permuted-blocks randomization procedure (Beller, Gecsk, & Keech, 2002) to ensure that the number of different types of foster parent–foster child pairs assigned to the experimental and control conditions would be as equal as possible. These types consisted of one foster parent paired, respectively, with one female foster child, one male foster child, one female and one male foster child, two female foster children, or two male foster children. In a single instance, two foster parents were paired with three male foster children, two to be tutored by one foster parent and the third by the second foster parent. This randomization procedure ensured that the number of parent–child pairs in the experimental or control conditions would be as equal as possible and that two or more children in a given foster home would be randomly assigned to the same condition (thereby eliminating the risk of leakage from one condition to the other and also making the tutoring in such cases logistically feasible).

### 2.3. Experimental conditions

#### 2.3.1. Wait-list control condition

The control children received the TYCW tutoring intervention during the school year (2009–2010) following that in which the experimental children had been tutored (2008–2009). During both years, each of the foster children in the experimental and control groups received a Registered Education Savings Plan (RESP) from their respective CAS for future postsecondary educational purposes. (RESPs are financial instruments created by the Government of Canada to encourage families and organizations such as CASs to save for children's post-secondary education.) Each child was assured of having $1400 deposited in his or her RESP account. The foster parents in the two groups agreed to communicate weekly or more often to their tutees that the RESP was a symbol of their value as persons and a concrete financial investment in their futures.

#### 2.3.2. Experimental condition

During 2008–2009, the foster children in the experimental group received tutoring and the RESP. The TYCW tutoring intervention was designed to provide 3 h per week of individual tutoring, for 30 weeks. The 3 h of weekly tutoring was to consist of 2 h of one-on-one direct instruction to the foster child in reading, 30 min of reading aloud by the foster child to the tutor or another adult in the home, and 30 min of self-paced instruction in math for the foster child, under the supervision of the foster parent. The math component was taught through step-by-step instruction in the form of a computer-based CD-ROM that the foster child used at his or her own pace. The reading component consisted of a four-level learn-to-read series of books, written by the designer of the TYCW program, Michael Maloney, and his team. For each reading level, there was a detailed instructor's manual and a student reader, and, for some levels, a student workbook as well.

To determine the level of the TYCW program at which the foster child was to begin, his or her current reading level was determined by means of a standard assessment passage, administered by a research team member immediately after the child had been randomly assigned to the tutoring or wait-list control group. Also, to promote behavioral self-regulation and optimal learning, the TYCW program incorporated a behavior-management component, based on a reward system in which the child was to be awarded points for positive behavior in a particular tutoring session.

### 2.4. Instruments

#### 2.4.1. Wide Range Achievement Test—Fourth Edition (WRAT4)

The WRAT4 (Wilkinson & Robertson, 2006), our outcome measure of the foster children's academic achievement, is a standardized, norm-referenced test that assesses basic reading and math skills. It was developed for use with individuals aged 5–94 or in Grades K–12. We used the Blue (rather than Green) Form at both the pre-test (September–October, 2008) and post-test (June, 2009), with all raw scores converted to standard scores (\( M = 100, SD = 15 \)). The psychometric properties of the WRAT4 Blue Form are excellent, with subtest internal consistency coefficients between .87 and .96 and good evidence of validity (Wilkinson & Robertson, 2006). The WRAT4 comprises four subtests, Word Reading, Sentence Comprehension, Spelling, and Math Computation, and also yields a Reading Composite score that is obtained by combining the Word Reading and Sentence Comprehension standard scores.

#### 2.4.2. Foster parent questionnaire

This questionnaire, developed by the study research team, was administered at the pre-test and post-test. The version for the experimental group asked the foster parents in the tutoring condition to report the number of lessons completed with their foster children during the intervention year in reading, the number of weeks spent on reading and math tutoring, the average amount of time spent weekly on different components of the tutoring program, and their use of the tutoring supports made available to them during the school year.

### 2.5. Procedure

#### 2.5.1. Training and coaching of foster parents

About a week before the pre-test assessment, the foster parents in the experimental group attended a 6-hour training session, conducted by the author of the TYCW program, Michael Maloney. No measure was administered to assess the foster parents' mastery of the tutoring intervention. The foster parents collected performance data during each tutoring session (e.g., the number of words read per minute by the foster child) and then forwarded the data weekly by email or telephone to the project coordinator. The latter sent these data to Mr. Maloney, who monitored each child's progress.
The foster parents were also encouraged to contact Mr. Maloney via email or a toll-free hotline for support and troubleshooting if they were experiencing difficulties in the implementation of the program or if a particular young person’s performance data indicated that he or she was not making adequate progress. Mr. Maloney and his staff also provided ongoing consultation, in the form of monthly telephone seminars, on-one-one coaching at the foster parent’s request by means of his toll-free tutor hotline, and project newsletters. The project coordinator was also in frequent contact with the foster parents by telephone or e-mail, to gather data and furnish encouragement.

2.5.2. Performance monitoring by foster parents

In monitoring the foster child’s reading skills, the foster parent measured sound fluency (i.e., the number of sounds read from a list of sounds per 30 s), word fluency (i.e., the number of words read from a word list per 30 s), and story fluency (i.e., the number of words read from a story in 1 min). The foster parent also recorded the number of lessons completed during the week and the amount of time spent on reading practice, during which the child read aloud to the foster parent or another adult in the home for practice. With regard to math, the foster parent collected data on the number of lessons completed by the child on the self-paced, computer-based math software, in the areas of addition, subtraction, multiplication, division, fractions, etc. Once submitted to the project coordinator, these data were placed into weekly spreadsheets and sent to Mr. Maloney for monitoring. If the project coordinator had not received data from a given foster parent in a timely way, she followed up with e-mail reminders and, as needed, telephone calls. The process of collecting these weekly data also served to assess the degree of fidelity with which individual foster parents were implementing the tutoring intervention.

2.5.3. Assessment of level of treatment fidelity and tutoring dosage

Judgments by the research team about the level of fidelity with which the foster parents in the experimental group had implemented the TYCW tutoring program (i.e., with high, medium, low, or zero fidelity, which in the context of this study also corresponded to a high, medium, or low level of tutoring “dosage”) were based on the weekly performance data sent to the project coordinator by the foster parents and, at the post-test, the questionnaire on which the foster parents recorded how many lessons they had actually covered during the year, the number of weeks of actual tutoring, and the average number of hours per week actually dedicated to tutoring. A few weeks after the post-test had taken place, the project coordinator and principal investigator went through the file of each child in the experimental group. A child was categorized as having received a high level of implementation and dosage of tutoring in reading if he or she had received a substantial number of weeks of tutoring (i.e., approximately 25 weeks or more), had completed a substantial number of lessons (i.e., approximately 60 lessons or more, or 30 lessons for young children, in the TYCW reading books), and had spent a substantial amount of time on the different components of the tutoring process per week. A rating of low implementation/dosage was made when the child had received only a relatively small amount of tutoring but still enough, in terms of weeks (i.e., more than 5 weeks), lessons (i.e., more than 10 lessons), and time per week to qualify as having received a low level of the intervention (rather than none at all). A child was categorized as having received a medium level of implementation/dosage when the degree of tutoring received was judged as less than high but more than low. Finally, children who were judged as having received no tutoring at all had clearly fallen below the low implementation/dosage threshold just defined. A similar procedure was used to assess implementation of tutoring in math.

2.6. Data analysis

2.6.1. ANCOVA via multiple regression

Gliner, Morgan, and Harmon (2003) noted that the analysis of covariance (ANCOVA) conducted through multiple regression has been recommended as the most powerful approach for analyzing pre-test/post-test comparison group designs such as ours. The ANCOVA approach takes account of any differences in the pre-test scores of the experimental and control groups, thereby reducing error variance by adjusting the post-test scores. Before carrying out the ANCOVAs, any missing data on the WRAT4 were calculated and imputed according to the instructions in the test manual. The foster children’s pre-test scores served as a single covariate, and their adjusted post-test scores were calculated by means of the procedure described by Huitema (1980, pp. 65–66). We tested the directional hypothesis (with a 1-tailed t-test of differences in means) that the foster children who had received tutoring would have higher adjusted mean scores at the post-test on the WRAT4 sub-tests than would the foster children in the wait-list control group.

2.6.2. Effect size index: Hedges’ g

As recommended by the What Works Clearinghouse (WWC) Procedures and Standards Handbook (WWC, 2008, pp. 35–37), we used the unbiased form (corrected for small sample size) of Hedges’ g as our index of the effect size (i.e., the standardized mean difference, or SMD). Hedges’ g is very similar in magnitude to but slightly more conservative than Cohen’s d, and the SMD is the unbiased difference between the mean outcomes of the intervention and control groups, divided by the pooled within-group standard deviation (SD). As is appropriate in computing Hedges’ g in the case of ANCOVA (see WWC, 2008, p. 37), we divided the adjusted difference between the post-test means of the tutoring and control groups by the unadjusted pooled within-group SD.

2.6.3. Improvement index

To show more clearly the practical impact of the tutoring intervention, we translated Hedges’ g into an “improvement index” (see WWC, 2008, p. 24). This was calculated as the difference between the percentile rank corresponding to the tutoring group mean and the percentile rank corresponding to the control group mean (by definition, the 50.0th percentile).

3. Results

3.1. Pre-test equivalence of experimental and control groups, and attrition

At the pre-test, there were no statistically significant differences (p > .05) between the experimental and control group foster children on gender, age, or the WRAT4 subtest means. Between the pre-test (September–October, 2008) and the post-test (June, 2009), 12 of the 42 foster children (28.6%) in the experimental group and 1 of the 35 (2.9%) in the wait-list control group withdrew from the study. This attrition, however, did not destroy the pre-test equivalence of the two groups, as the pre-test means of the 30 foster children remaining at the post-test in the experimental group (17 girls, 13 boys) and the 34 in the wait-list control group (19 girls, 15 boys) were not significantly different (p > .05) on gender, age, or the WRAT sub-tests.

From the numerous contacts by telephone and e-mail that the study coordinator had had with the foster parents during the intervention year, we were able to classify the reasons for attrition as either mainly endogenous (directly related to either the intervention or control condition), mainly exogenous (unrelated to either condition), or mixed (both). In the experimental group, 5 of the 12 cases of attrition were mainly endogenous, 5 were mainly exogenous, and
2 were mixed. In the wait-list control group, the single case of attrition was exogenous, caused by a change in placement. The foster parents in the experimental group cited several endogenous reasons for dropping out of the study: the tutoring took much time for very busy caregivers, was a source of conflict between the tutor and child, or was not needed because the foster child was already doing well in school. The exogenous reasons that the foster parents mentioned included illness on their part or that of the foster children who were being tutored, a change in children’s placements, or children’s arriving too late at the post-test for valid administration of the post-test measures. The mixed category consisted of a combination of both types of reasons.

3.2. Effects of the TYCW tutoring intervention

Table 1 shows the means on the pre-test scores and on the ANCOVA-adjusted post-test scores for the experimental (tutoring) and wait-list control groups, on the five WRAT4 sub-tests. On Word Reading, which assesses the decoding of letters and words through identification and word recognition (Wilkinson & Robertson, 2006), the difference between the adjusted post-test group means was not statistically significant (t(62) = 0.90, p = .39, 1-tailed), and the size of the effect (Hedges’ g = 0.19) was below the 0.25 threshold that WWC (2008) considers “substantively important”. The improvement index was 3.6% (i.e., 53.6th–50.0th percentiles), such that the average foster child in the tutoring group was at the 53.6th percentile of the control group, whose average foster child was, by definition, at the 50.0th percentile.

On Sentence Comprehension, measuring the ability to gain meaning from words and comprehend ideas and information contained in sentences, the difference between the adjusted post-test group means was statistically significant, in the hypothesized direction (t(62) = 1.85, p = .035, 1-tailed). The effect size (g) was 0.38, well above the WCC (2008) threshold of substantive importance, and the improvement index was 14.8% (64.8th–50.0th percentile).

On Reading Composite, obtained by combining the standard scores for Word Reading and Sentence Comprehension, the statistical significance of the difference between the adjusted post-test group means was at the level of a trend (t(62) = 1.32, p = .096, 1-tailed). The effect size (g) was 0.29, above the WCC (2008) threshold of 0.25, and the improvement index was 10.4% (60.4th–50.0th percentile).

On Spelling, which assesses the ability to encode sounds into written form through the use of a dictated spelling format that contains percentile), the effect size (g) was noted, however, that the mean effect for math in Harper and Schmidt’s randomized group-based study (2012) of Maloney’s (1998) TYCW tutoring program, which they conducted with a mainly Aboriginal sample of children in care of primary-school age, our significant results on the WRAT4 were of generally similar size but were found on different subtests. Specifically, on Word Reading, the effect that Harper and Schmidt found (Hedges’ g = 0.42, p = .01) was statistically and practically significant, whereas ours was not (g = 0.19, ns); on Sentence Comprehension, their effect (g = .095, ns) was very small and non-significant, whereas ours was both statistically and practically significant (g = 0.38, p < .05); on Spelling, their effect (g = 0.38, p < .01) was statistically and practically significant, whereas ours was not (g = −0.08, ns); and on Math Computation, their effect (g = 0.26, ns) was statistically non-significant but substantively important, whereas ours was significant in both senses (g = 0.46, p < .01). As Harper and Schmidt (2012) remarked, these differences may be due to several factors, including their group tutoring format versus our individualized format or differences in the way that our respective studies implemented certain aspects of the reading and mathematics components of the TYCW tutoring model. What seems most important to underline in the present state of knowledge are two broader and

Finally, on Math Computation, measuring the ability to perform basic mathematical computations through counting, identifying numbers, solving simple oral problems, and calculating written math problems, the difference between the adjusted post-test means was statistically significant, in the hypothesized direction (t(62) = 2.43, p = .009, 1-tailed). The effect (g) of tutoring was 0.46, and the improvement index was 17.7% (67.7th–50.0th percentile).

4. Discussion

We tested the hypothesis that foster children in primary school who were exposed to individualized direct-instruction tutoring by their foster parents would experience statistically and practically greater pre-test to post-test gains in the basic academic skills of reading and math than would foster children in a wait-list control group. Overall, our results were supportive of our hypothesis, but only in part. First, with respect to the improvement of reading skills, tutoring had a statistically and practically significant impact on Sentence Comprehension and a substantively important effect on Reading Composite that was statistically significant at the level of a trend. Second, with respect to gains in math skills, tutoring produced a statistically and practically significant gain on Math Computation. Third, on two reading-related measures, Word Reading and Spelling, tutoring had an effect that was neither statistically nor practically significant.

The effects of tutoring in our study, on the whole, were in the same range as those reported by Ritter et al. (2009) in their meta-analysis of randomized studies of tutoring in children of primary-school age in the general population, although some of our effects tended to be of different magnitudes in comparable academic domains. On Reading Composite, the effect of tutoring in our study (Hedges’ g = 0.29, statistically significant at the level of a trend) was similar to the statistically significant mean effect found by Ritter et al. (2009) on Reading Global (Hedges’ g = 0.26). On the other hand, the statistically significant impact of tutoring that we found on Sentence Comprehension (g = 0.38) was considerably larger than the non-significant mean effect (g = 0.18) found by Ritter et al. (2009) on Reading Comprehension. Conversely, our non-significant effect on Word Reading (g = 0.19) was clearly smaller than the statistically significant mean effect for Reading Words and Letters (g = 0.41) reported in their meta-analysis. Finally, our statistically significant and strongest effect, on Math Computation (g = 0.46), was larger than their non-significant mean effect (g = 0.26). (It should be noted, however, that the mean effect for math in Harper and Schmidt, 2009, was based on only 5 studies and was above the WCC (2008) threshold of g = 0.25 for being considered substantively important.)

Compared to the results of Harper and Schmidt’s randomized group-based study (2012) of Maloney’s (1998) TYCW tutoring program, which they conducted with a mainly Aboriginal sample of children in care of primary-school age, our significant results on the WRAT4 were of generally similar size but were found on different subtests. Specifically, on Word Reading, the effect that Harper and Schmidt found (Hedges’ g = 0.42, p < .01) was statistically and practically significant, whereas ours was not (g = 0.19, ns); on Sentence Comprehension, their effect (g = .095, ns) was very small and non-significant, whereas ours was both statistically and practically significant (g = 0.38, p < .05); on Spelling, their effect (g = 0.38, p < .01) was statistically and practically significant, whereas ours was not (g = −0.08, ns); and on Math Computation, their effect (g = 0.26, ns) was statistically non-significant but substantively important, whereas ours was significant in both senses (g = 0.46, p < .01). As Harper and Schmidt (2012) remarked, these differences may be due to several factors, including their group tutoring format versus our individualized format or differences in the way that our respective studies implemented certain aspects of the reading and mathematics components of the TYCW tutoring model. What seems most important to underline in the present state of knowledge are two broader and

| Table 1 Mean of the pre-test scores and ANCOVA-adjusted post-test scores for the experimental (tutoring) and wait-list control groups, on the WRAT4 sub-tests. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| WRAT4 sub-test  | WL control group| Experimental group|                 |
|                 | (n = 34)         | (n = 30)         | Pre-test        | Post-test       | Pre-test        | Post-test       |
| Word reading    | 97.21            | 97.78            | 96.80           | 100.32          |
| Sentence comp.  | 96.21            | 98.69            | 97.67           | 103.22          |
| Reading comp.   | 95.65            | 97.44            | 96.37           | 101.23          |
| Spelling        | 98.53            | 98.87            | 95.53           | 97.67           |
| Math comp.      | 84.85            | 86.30            | 87.10           | 92.10           |

Note. WL = Wait-list. The post-test scores were adjusted via ANCOVA for differences on the pre-test scores.
encouraging conclusions reached by Forsman and Vinnerljung in their review (2012), namely, that most educational interventions with children in care appear to produce positive results and that replications of tutoring interventions, including the TYCW approach evaluated by us and by Harper and Schmidt, can be undertaken with grounds for optimism.

The fact that tutoring by foster parents has now been shown to improve foster children’s basic academic skills is important for at least two reasons. First, it has the potential to enable child welfare services to capitalize on the potential of a readily available but relatively untapped resource—foster parents—to improve educational outcomes. The implementation of such a strategy would be consistent with Jackson’s (2007) strong advocacy in favor of a much greater role for foster parents in the academic education of their foster children. Second, in being based in the foster home rather than in the school, tutoring by foster parents would complement current school-based efforts (e.g., Tideman, Vinnerljung, Hintze, & Isaksen, 2011) to improve educational performance. In Ontario, tutoring by foster parents would improve foster children’s academic skills, thereby enabling them to take maximum advantage of the system-change efforts now taking place in the province. These systems-level initiatives have taken the form of Crown Ward Education Championship Teams, in which young people in long-term care receive collaborative mentoring, peer support, and guidance from a range of actors in child welfare agencies, schools, colleges, and universities (Ontario Association of Children’s Aid Societies, 2007).

Regarding treatment fidelity and the dosage of tutoring, we verified at the post-test that none of the 34 children in the wait-list control children had received the TYCW tutoring program during the intervention year (the control children received the TYCW program in the following school year). After the post-test, two members of the research team examined together and discussed the files of each of the 30 foster children in the experimental group. The team members rated 21 (70%) as having experienced a high level of treatment fidelity in reading, 2 (7%) as having had a medium level of fidelity, and 7 (23%) as having received a low level. Overall, the number of TYCW reading lessons that had been completed varied between 10 (very low) and 128 (very high), with the average (M = 63.6) quite close to the intended number of 60. In math, 17 (57%) of the experimental-group children were rated as having experienced a high level of treatment fidelity, 7 (23%) a medium level, 5 (17%) a low level, and 1 (3%) little or no fidelity at all. To check the reliability of our rating of treatment fidelity, approximately one year after the initial fidelity ratings had been made, three members of the research team (including the two who had made the initial assessment) again went through the files of each of the 30 children in the experimental group and re-rated them (without knowledge or recall of the child’s initial rating). In 29 out of 30 cases (97%), the re-rated level of treatment fidelity in reading was identical, and in the other case (3%), it differed by only one level. No attempt was made to re-assess the initial assessment of treatment fidelity in math tutoring because the foster parent’s role in supervising the child’s self-paced, computer-based work in math was considerably looser and more informal than in reading. As a result, the research team found treatment fidelity in math tutoring to be much harder to judge reliably.

With regard to the key implementation-related issue of acceptability of the TYCW tutoring program to various audiences, a small number of CAS managers and staff voiced objections, during our early efforts to recruit CASs, that our use of a randomized design would be unfair to those foster children eventually assigned to the control group and hence even unethical. We responded in several ways. First, we explained the importance of using the strongest possible (i.e., randomized) research design, to arrive at maximally robust conclusions about the effectiveness of the TYCW tutoring program, an especially important consideration given the few evidence-based interventions that have been shown to be effective in improving foster children’s educational achievement. Also, other CAS staff members, who had backgrounds in research methods, argued in favor of the desirability of a randomized design. Second, we made it clear that all of the children in the wait-list control group would be offered the integral TYCW intervention during the following school year. Third, we emphasized that our tutoring intervention was being offered over and above standard care, such that the foster children in the control group would not be deprived of any services that they would otherwise receive.

Concerning the acceptability of the tutoring program to the foster parents in the experimental group, the majority had a favorable perception of it: 79% said they would recommend the TYCW program “without hesitation” to other foster parents, 14% said they would recommend the program “with some hesitation”, and only 7% would not recommend it. This positive view existed despite a number of obstacles identified by the foster parents in the experimental group, including the tutored child’s behavior problems (mentioned by 57% of the foster parents); the time demands of the foster household’s involvement in extracurricular activities (50%); the resistance of the foster child to the tutoring (40%); the amount of other homework that the child had to do (40%); and foster–parent illness lasting longer than two weeks (23%). Concerning the requirement of spending 3 h a week on tutoring, 38% of the foster parents in the experimental group said that this was “a difficult goal to attain, but I managed”, and 35% answered that it was “reasonable, and I was able to achieve it”. On the other hand, 24% stated that it was “too much time to ask of foster parents”, and 3% felt that the requirement was “reasonable, but I was unable to achieve it at least half the time”.

Despite these obstacles, the project coordinator noted, in her frequent contacts via e-mail and telephone with the foster parents, that many in the experimental group had positive things to say about the tutoring program, often attributing it to desirable changes in the foster children. Some of these anecdotal observations were as follows (identifying information has been removed to safeguard participants’ anonymity):

- “The foster child is doing very well. She has reached 200 words/min in fluency, which brought tears to her and her foster parent’s eyes. The child is excited because she sees her progress and thanks her foster parent every night.”
- “The foster child reports that her reading is much better in class and that she is participating more in the regular classroom.”
- “The foster child is enjoying the one-on-one special time he gets during the tutoring.”
- “The foster child’s interest in the tutoring is unbelievable. The child’s sentences and words are getting longer, and pronunciation is a lot better.”
- “The foster child knows he’s getting smarter as he learns more words and cooperates well with the tutoring.”
- “The foster child is doing very well. She really loves the math and can’t wait until she catches up to her class!”

With respect to the perceived quality of the tutoring training, most of the foster parents in the experimental group said at the post-test that they believed they had been adequately trained to tutor during their pre-intervention six-hour, one-day training session. They rated the TYCW reading-related materials and training procedures much more favorably, however, than the math-related or behavior-management materials and training component. This no doubt reflected the fact that the materials and training procedures dealing with reading were much more comprehensive, with a detailed instructor’s manual and accompanying student manual, than the math or behavior-management elements. The foster parents’ reactions, including their identification of foster children’s behavior problems as the single most frequently encountered obstacle to tutoring, suggest that the training materials and procedures for math tutoring and behavior-management should be improved. In practice, this would no doubt require a two-day rather than one-day training
session. Despite these perceived weaknesses, it is interesting that we found that tutoring produced its largest effect on Math Computation. This is perhaps not so surprising when we recall that our pre-test results had shown (in line with very similar findings reported by Trout et al., 2008) that math was the academic-skill area in which, on average, the foster children in our sample were weakest.

Regarding future intervention studies on tutoring for young people in care, it would be desirable to recruit a larger number of local child welfare organizations and, concomitantly, a larger sample of foster children and foster parents. Now that we and Harper and Schmidt (2012) have shown that the TYCW tutoring produces positive effects for foster children in care of about the same magnitude as those found by Ritter et al. (2009) in tutoring programs for children in the general population, it should be easier to interest a wider range of agencies in efforts to replicate and extend our results. Any attempts to implement our intervention on a wider, “scaled-up” basis should note that of the 42 foster children originally randomized to the experimental group, 50% received TYCW tutoring at a high level of fidelity, 21% received it at medium or low levels, and 29% received little or no tutoring at all. Based on an admittedly small sample that, furthermore, was located in nine far-flung sites, our findings suggest that many foster parents and children would be able to carry through on a commitment to implement the program. This could potentially involve thousands of foster children and foster parents in a jurisdiction such as Ontario. At the same time, our findings suggest that the TYCW program is not suited to everyone, and strong efforts should be made to engage only those who are motivated and able to use it well.

Overall, it is encouraging that we found it possible to improve foster children’s educational achievement through tutoring, as improvements in basic academic skills in primary school are likely to make it easier to cope successfully with the educational demands of high school. Also, the finding by Berlin et al. (2011) in Sweden that success in primary school may help prevent the development of serious psychosocial problems in young adulthood provides an added reason to make educational improvement a high research, practice, and policy priority in child welfare. In evaluating the effects of promising interventions, it will be important, as advocated by Forsman and Vinerljung (2012), to increase the number of studies and to use large samples and experimental designs. This will allow the incremental construction of a growing base of credible interventions that child welfare personnel, foster parents, and foster children may employ with confidence.

References