

Does What Works Clearinghouse Work? A Brief Review of Fast ForWord®

Genevieve McArthur*

Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia

The What Works Clearinghouse (WWC) provides online reports to the public about the scientific evidence for educational interventions. The quality of these reports is important because they effectively tell the non-scientific community which programmes do and do not work. The aim of this brief review is to assess WWC's report on a clinically popular, yet theoretically controversial, intervention called Fast ForWord® (FFW). Some of the methods used by WWC to assess FFW were problematic: the literature review included studies that had not passed peer review; it failed to include a key study that had passed peer review; alphabetic skills were assessed with phonological awareness outcomes; effectiveness ratings were based on statistical significance; terms peculiar to WWC were not clearly defined; and existing quality control procedures failed to detect an error in the WWC report. These problems could be addressed by making minor adjustments to WWC's existing methods and by subjecting WWC reports to the scientific peer-review process before they are released to the public.

In 2002, the US Department of Education's Institute of Education Sciences developed the What Works Clearinghouse (WWC; see <http://ies.ed.gov/ncee/wwc/>) to provide the public with reports about the scientific evidence for the effectiveness of educational interventions. These reports are important because they provide a much needed, and all too rare, interface between scientific research and the public. They are also potentially powerful because they effectively tell the non-scientific community which educational programmes work and which do not. For this reason, it is imperative that WWC produces valid assessments about the scientific evidence for an intervention. The aim of this brief review is to assess WWC's report on a clinically popular, yet theoretically controversial (Cohen et al., 2005), education intervention called Fast ForWord® (FFW). After briefly describing FFW, this review will summarise the stages involved in the WWC report process, assess the methods involved in each of these stages, and then suggest how some of these methods might be improved.

FFW

FFW comprises a suite of software programmes (*ForWord® Language*, *Fast ForWord® to Reading Prep*, *Fast ForWord® to Reading 1*, and *Fast ForWord® to Reading 2*)

*Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia.
Email: genevieve.mcarthur@mq.edu.au

developed by Scientific Learning Corporation (2007a). This suite is designed to help children's reading and spoken language by developing their memory, attention, processing, and sequencing using computer-based exercises that train their listening accuracy, phonological awareness, and language structures. Clients train for 30 to 100 minutes per day, 5 days per week, for 4 to 16 weeks. Although this suite is designed to improve spoken language skills as well as reading, the report reviewed here (dated July 9, 2007) only assessed FFW's effect on alphabets and reading comprehension (Institute of Education Sciences, 2007a).

Stages in the WWC Review Process

According to the WWC site (Institute of Education Sciences, 2007b) WWC invites suggestions for an intervention from the public and experts. Once an intervention has been selected, a *review team* is formed comprising a Principal Investigator (one of WWC's eight researchers who are experts in their own field of research), a Project Coordinator (an expert in general methodology), and research analysts (trained in reviewing and summarising scientific evidence). This team conducts a *literature review* of published and unpublished studies of an intervention from various sources including scientific journals, the Internet, the intervention's developers, and the general public. The team categorises each study according to WWC Evidence Standards: 'Meets Evidence Standards' (randomised controlled trials and regression discontinuity studies), 'Meets Evidence Standards with Reservations' (randomised controlled trials and regression discontinuity studies with flawed randomisation, attrition, or disruption) and 'Does Not Meet Evidence Screens'. Studies that meet evidence standards (with or without reservations) are selected for the *final sample*. The *methods* and *outcome measures* of the final sample of studies are summarised in tables. The *statistics* of the final sample of studies are converted into comparable measures of effect size and statistical significance. These measures are used to *rate the effectiveness* of the intervention on different domains (e.g., reading comprehension and 'alphabets' for reading interventions). WWC also calculates the '*extent of evidence*' (based on the number of studies, schools, and subjects) for each domain of interest. An Overview of the intervention report is created which comprises a summary of the intervention, the 'extent of evidence' rating, the effectiveness rating, and links to supporting information. This Overview is published on the WWC website as the homepage for the intervention report.

Assessing the Methods Used at Each Stage of the WWC Review Process

The Review Team

WWC did not specify who was on the review team for the FFW intervention report. However, the WWC site does provide information about the WWC's Principal Investigators (Institute of Education Sciences, 2007c). Two of these people are

experts in the theory of reading but have not published in the field of research associated with FFW.

The Literature Review

The WWC team reviewed 24 published and 91 unpublished documents (Institute of Education Sciences, 2007d). Unfortunately, this literature review did not include a study by Cohen et al. (2005), which is one of only a few FFW studies that includes all the necessary controls. It is not clear why this study was excluded. It cannot be because Cohen et al. included inappropriate reading measures or because they only included spoken language measures because WWC included 13 other studies that they footnoted as having unsuitable reading tests and two studies that only included spoken language measures (Merzenich et al., 1996; Tallal et al., 1996). In line with Carter and Wheldall's comment in the editorial (2008), the absence of Cohen et al. from the WWC literature review suggests that being an expert in methodology does not guarantee a thorough review of the scientific literature related to an intervention.

The Final Sample

Only six studies passed WWC Evidence Standards (Institute of Education Sciences, 2007d). Four of these studies were conducted by Scientific Learning Corporation—the developers of FFW—and were published in their own *MAPS for Learning: Educator Reports* or *MAPS for Learning: Product Reports* (Scientific Learning Corporation, 2007b). None of these studies was critically assessed by independent experts in the field of research that relates to FFW (i.e., none had passed peer review). The remaining two studies were carried out by independent researchers but had not passed peer review. If WWC wants to help the public choose scientifically proven interventions then it needs to assess the existing data scientifically. The scientific method demands that a study passes peer review before it is released to the public. WWC does not adhere to this method.

Methods

The tables that summarised the methods of each study that met WWC Evidence Standards made it easy for the reader to understand the techniques used in each study (see Institute of Education Sciences, 2007e).

Outcome Measures

The tables that summarised the outcome measures provided a clear description of the tests used to assess the efficacy of FFW (Institute of Education Sciences, 2007e). However, phonological awareness tasks were included as outcome measures of alphabetic skills. Alphabetic skills depend on the ability to recognise letters and

associate these letters with their respective speech sounds. Phonological awareness depends on the ability to identify and manipulate speech sounds. Phonological awareness does not involve letters and so is *not* an alphabetic skill. Until recently, it has generally been assumed that phonological awareness has a causal effect on alphabetic skills. However, a careful review of the literature by Castles and Coltheart (2004) revealed that this assumption has not yet been proven to be reliable. Thus, WWC should not use phonological awareness outcomes when assessing the effectiveness of FFW on alphabetic skills. Most experts in the theory of reading development would be aware of this issue.

Statistics

WWC have clearly put a lot of time and effort into creating a set of procedures that convert the different statistics reported by different studies into common measures of effect size and statistical significance. These measures were outlined in tables which were useful (Institute of Education Sciences, 2007e). However, it was not easy to find a definition of a ‘substantively important’ effect size. This was buried in a PDF called Technical Details of WWC-Conducted Computations (Institute of Education Sciences, 2007f). A clear definition of ‘substantively important’ effect sizes needs to be included with the tables.

Effectiveness Ratings

WWC bases its effectiveness ratings on the number of studies showing statistically significant or ‘substantively important’ effects. As WWC point out themselves in yet another technical paper (Institute of Education Sciences, 2007g), statistical significance testing can be misleading: Even small effects can be statistically significant given a large enough sample. This means that WWC would award an intervention tested with a large sample a higher rating than an intervention tested with a small sample even if both interventions had exactly the same effect. This would not be a problem if the ratings were based on measures of effect size which do not depend on sample size.

WWC rates an intervention’s effect on a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. In the domain of alphabets, WWC rated *Fast ForWord*[®] as having *mixed effects* because ‘It did not meet the criteria for positive effects because no studies showed statistically significant positive effects’ and ‘it did not meet the criteria for potentially positive effects because two studies showed indeterminate effects and only one study showed a substantively important positive effect’ (Institute of Education Sciences, 2007e). This was puzzling because the statistics showed that two studies found statistically significant effects for alphabetic skills. Indeed, WWC subsequently stated that (1) two studies of *Fast ForWord*[®] had a statistically significant positive effect in this domain, and both had strong designs, and (2) no studies of *Fast*

ForWord[®] showed statistically significant or substantively important negative effects in this domain. Thus the effect of FFW on alphabetic skills actually met the criteria for a positive effect rather than a mixed effect as the rating stated.

A possible explanation for this inconsistency emerged in the subsequent paragraph on the ratings for the reading comprehension domain. Here too WWC stated that *Fast ForWord*[®] had *mixed effects* because ‘It did not meet the criteria for positive effects because no studies showed statistically significant positive effects’ and ‘it did not meet the criteria for potentially positive effects because two studies showed indeterminate effects and only one study showed a substantively important positive effect’. This time the rating matched the statistics as well as the subsequent statement that ‘one study had substantively important and positive effects and one study had substantively important and negative effects’. Perhaps the true rating of FFW’s effect on the *alphabetic domain* (i.e., positive effect) had been replaced with a copy of FFW’s effect on the *reading comprehension* domain (i.e., mixed effect). This error demonstrates that the quality control procedures that WWC have in place could be improved. According to the WWC website (Institute of Education Sciences, 2007b) these involve an ‘opportunity’ for the study’s authors to review the coding results, an ‘opportunity’ for the developer of the intervention to review the intervention report, and a review by the WWC Steering Committee. Note that these quality control procedures do not include independent peer review of the intervention report.

Extent of Evidence

WWC also provide ratings for the ‘extent of evidence by domain’ (Institute of Education Sciences, 2007e). These ratings depend upon the number of studies, the number of schools, and the size of the samples used in the reviewed studies. This is a useful measure of power.

The Overview

Fortunately, the error made on FFW’s effectiveness rating on the alphabetic domain was not carried over to the Overview page since this is the homepage (i.e., first point of contact) for WWC’s intervention report on FFW (Institute of Education Sciences, 2007a). The only problem with this page is the use of the term ‘extent of evidence’. This term—a measure based on the number of studies, the number of schools, and the size of the samples used in the reviewed studies (i.e., amount of data)—is defined in a footnote under the ‘extent of evidence’ table at the end of the appendices. Many readers would not read this footnote before they read the Overview page. It is easy to interpret ‘extent of evidence’ as ‘strength of evidence’. So when the Overview page states ‘the extent of evidence for *Fast ForWord*[®] to be small for alphabetics and comprehension’ a reader may form the impression that FFW has only a small effect on these domains. They may subsequently be confused to read that ‘*Fast ForWord*[®] was found to have positive effects on alphabetics and mixed effects on

comprehension'. The use of a less ambiguous term, or a definition on the Overview page of what 'extent of evidence' actually means, would be helpful.

Possible Improvements in WWC Methods

The aim of this brief review is to assess WWC's report on a clinically popular, yet theoretically controversial, intervention called Fast ForWord® (FFW). This review identified methodological problems with the WWC process that question the validity of its ratings: the literature review included studies that had not passed peer review; it failed to include a key study that had passed peer review; alphabetic skills were assessed with phonological awareness outcomes; effectiveness ratings were based on statistical significance; terms peculiar to WWC were not clearly defined; and existing quality control procedures failed to detect an error in the WWC report. Most of these problems could be fixed with some minor adjustments in methodology. Specifically, the research review could only include peer-reviewed studies; phonological awareness could be removed as a measure of alphabetic skill; the effectiveness ratings could be based on effects sizes; WWC-specific terms such as 'extent of evidence' and 'substantively important' could be more clearly defined; and quality control procedures could be adjusted to more reliably detect errors in reports.

A more significant problem with the WWC review process is that it does not include experts in the theory associated with an intervention under scrutiny. Clearly, WWC cannot employ experts in every field of research in education. However, WWC's intervention reports could be subjected to the same rigorous peer-review process that other scientific meta-analyses undergo before they are accepted for publication. This could be done if WWC aligned itself with a high-impact journal in education or psychology. The editors of this journal could send WWC's intervention reports out to experts in the appropriate field of research. Once a WWC intervention report met the satisfaction of at least three such experts, it could be published on the WWC site. Subjecting WWC intervention reports to peer review would help WWC provide the public with valid judgements about which educational interventions do and do not work.

Acknowledgements

Genevieve McArthur is supported by an Australian National Health and Medical Research Council (NHMRC) Project grant (235305), an NHMRC Howard Florey Centennial Fellowship (235111), and by the Macquarie Centre for Cognitive Science which is an Australian Research Council Special Research Centre (S00001507). I would like to thank the editors and anonymous reviewers for their insightful comments and constructive criticisms.

References

- Carter, M., & Wheldall, K. (2008). Why can't a teacher be more like a scientist? Science, pseudoscience and the art of teaching. *Australasian Journal of Special Education*, 32(1), 5–21.

- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, 91, 77–111.
- Cohen, W., Hodson, A., O'Hare, A., Boyle, J., Durrani, T., & McCartney, E. et al. (2005). Effects of computer-based intervention through acoustically modified speech (Fast ForWord) in severe mixed receptive-expressive language impairment: outcomes from a randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, 48, 715–729.
- Institute of Education Sciences. (2007a). *Intervention Fast ForWord®: Overview*. Retrieved November 12, 2007 from http://ies.ed.gov/ncee/wwc/reports/beginning_reading/fastfw/index.asp.
- Institute of Education Sciences. (2007b). *Overview: Review process*. Retrieved November 12, 2007 from <http://ies.ed.gov/ncee/wwc/overview/review.asp>.
- Institute of Education Sciences. (2007c). *Overview: Principal investigators*. Retrieved November 12, 2007 from <http://ies.ed.gov/ncee/wwc/overview/index.asp?ag=pi>.
- Institute of Education Sciences. (2007d). *Intervention: Fast ForWord®: References*. Retrieved November 12, 2007 from http://ies.ed.gov/ncee/wwc/reports/beginning_reading/fastfw/references.asp.
- Institute of Education Sciences. (2007e). *Intervention: Fast ForWord®: Appendices*. Retrieved November 12, 2007 from http://ies.ed.gov/ncee/wwc/reports/beginning_reading/fastfw/appendix.asp.
- Institute of Education Sciences. (2007f). *Technical Details of WWC-Conducted Computations*. Retrieved November 12, 2007 from http://ies.ed.gov/ncee/wwc/pdf/conducted_computations.pdf.
- Institute of Education Sciences. (2007g). *Effect Size Substantive Interpretation Guidelines: Issues in the Interpretation of Effect Sizes*. Retrieved November 12, 2007 from <http://ies.ed.gov/ncee/wwc/pdf/essig.pdf>.
- Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science*, 271, 77–81.
- Scientific Learning Corporation. (2007a). *Fast ForWord Products*. Retrieved November 12, 2007 from <http://www.scilearn.com/products/index.php>.
- Scientific Learning Corporation. (2007b). *PDF Reports*. Retrieved November 12, 2007 from <http://www.scilearn.com/results/main=downloads>.
- Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., & Nagarajan, S. S. et al. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, 271, 81–84.

Copyright of Australasian Journal of Special Education is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.