

表紙

題名

ボランティアによる個別指導プログラムの有効性： 系統的レビュー

レビューワ

Ritter G, Denny G, Albin G, Barnett J, Blankenship V

日程

編集日：

実質的な最終更新日：

部分的な最終更新日：

次段階予定日：

プロトコル初版

レビュー初版：

レビューワ連絡窓口

ゲーリーW.リッター博士（Dr. Gary W. Ritter）

助教授、教育と社会政策

アーカンソー大学

207 Graduate Education Building Fayetteville, AR 72701

電話： (479) 575-4971

ファックス： (479) 575-3196

電子メール： garyr@uark.edu

URL： <http://policy.uark.edu/ritter>

レビューワの調査分担

GR（レビューワ代表）は本レビューの執筆を行い、GD は、GR の指導のもと、メタ分析（meta-analysis）を主導した。GA と JB は GR と共に検索方法の策定と実施の作業にあたった。GD は、GA と JB の助言を経て、すべての完全にコード化された論文からのデータの抽出作業にあたった。VB は GR と共に当初の論文検索とプロトコルの開発にあたった。

内部支援組織

アーカンソー大学

外部支援組織

2005 年における本レビューの完成作業は、キャンベル共同計画教育レビュー部会からの助成金を受けて行われた。それ以前の本プロジェクトに対する資金援助は、ペンシルバニア大学とミズリー大学向けの助成により、スミス・リチャードソン財団によってなされた。

先導的レビューのプロトコル草案は、レベッカ・メイナード博士(Dr. Rebecca Maynard)の率いるペ

ンシルバニア大学教育学大学院の系統的レビューとメタ分析の実施に関するセミナー参加者を含む多くの同僚の協力を得て完成にいたった。

ボランティアによる個別指導の分野への最初に検討されたのは、1998-2000年の間、ペンシルバニア大学の教授陣と職員が実施した“フィラデルフィア西部における個別指導プロジェクトの評価”のテーマの下に行われた。

系統的レビューは、現在アーカンソー大学が実施し、その財政支援も行っている。しかし、本調査の結果は必ずしもスミス・リチャードソン財団、ペンシルバニア大学及びアーカンソー大学の公式な見解や方針を表すものではない。

謝辞

本レビューの開始時から完成に至るまで監督と指導を頂いたレベッカ・メイナードに感謝する。本レビューに数々の有益な意見や提案を賜ったジェラルディン・マクドナルドとジェフ・ヴァレンティンに感謝する。マーク・リップシー、チャド・ナイ、ハープ・ターナー、及びロバート・ボルチからもらった技術的助言も大いに役立った。

潜在的利益相反

リッター調査 2000 を含めるべきか否かの決定にはゲーリー・リッターは参加しなかった。

日程

レビューのフォーマット変更日：

新たな調査を探したが見つからなかった日：

新たな調査が見つかったが含めるに至らなかった／除外扱いとなった日：

レビューワの結論のセクションが修正された日：

コメント／批評が追加された日：

コメント／批評に対する返答が追加された日：

レビューの本文

概要

学校の説明責任(accountability)を要求する声が高まるなか、学校の管理者は、個別指導などの生徒の学力を向上のための方法を探し求めている。現在用いられている個別指導法には、ボランティアによる個別指導、仲間による個別指導、縦割り(cross-age)の個別指導、マンツーマン個別指導等がある。近年ボランティアによる個別指導に関連するエビデンスがかなり多く入手できるようになった。従って、本レビューは、生徒が処遇群(treatment group)と対照群に無作為割当されたボランティアによるプログラムの評価だけに基づいている。我々は、ボランティアによる個別指導プログラムの有効性の評価に役立てるよう、無作為実地試験(randomized field trials)に関する報告を行っている 21 件の調査を見つけた(これらの調査のうちのいくつかは複数のコーホートに関する別個の報告を行っていたので、27 の調査コーホートである)。

要約

背景

ボランティアによる個別指導プログラムが意図するところは、生徒の成績を上げること、良き指導者を提供すること、ならびに生徒の自尊感情と行動様式を高めることである。大方の学校の現況は、熱心さと努力は非常になされているが、どのボランティアによる個別指導プログラムが最も効果的なのかに確信が持てないところにある。従って、我々がここに強く主張したいのは、ボランティアによる個別指導プログラムに関する既存の文献を徹底分析すると個別指導プログラムの最も効果的なタイプに関する情報を学校に提供できるという主張である。

目的

米国の幼稚園から 8 年生 (K-8)までのの学力を向上させるためのボランティアによる個別指導プログラムの有効性を査定し、さらに個別指導が効果的となるのはどのような生徒に対してで、かつどのような条件の下においてであるかを詳細に調査すること。

検索方法

以下の社会科学データベースによる検索を実施した。

①Academic Search Premier、②Primary Search、③Professional Development Collection、④Middle Search Plus、⑤Psychology and Behavioral Sciences Collection、⑥PsycINFO、⑦Sociological Collection、⑧ ERIC (Education Resources Information Center)、⑨Proquest Digital Dissertations

選択規準

1985 年 1 月から 2005 年 8 月に至る間に出版されたもののうち、無作為実地試験(特定の処遇群が処遇を受けなかった対照群と比較されているもの)のみが我々のレビューの対象となった。調査には他にも部分的に関係するもの(例えば、行動や感情に与える影響)があったが、我々は学力のみを対象とした。幼稚園から 8 年生 (K-8)向けのプログラムで、成人で正規ではない(ボランティアに

よるの)教員を使って実施されたもののみを対象とした。

データの収集と分析

様々なボランティアによる個別指導プログラムや様々な調査特性を持つものの成績が分析できるように、各調査の方法や手順（研究計画、分析、アウトカムの指標）、介入の仕方（期間、条件設定）ならびに対象者サンプル（年齢、性別、民族性）に関する詳細はコード化された。我々がレビュー対象に含めた調査について、測定された全アウトカムを特定した後、メタ分析用に、効果量(effect size)の適当な測定規準(metric)を選んだ。我々は、各アウトカム指標に標準化された標準化平均差（standardized mean difference:SMD）効果量統計のヘッジズ(Hedge)の不偏推定値 g （特定の結果変数に対する処遇群と対照群の平均値間の差をそれらの統合標準偏差(pooled standard deviations)で割った値）を用いた。平均値や標準偏差値の記述がないときは、我々は効果量の推定値をウィルソンとリップシー(Wilson and Lipsey, 2001)が推奨する手法を試用した。調査が一つのアウトカムに対し複数の尺度を掲げているときは、我々は効果量の統合平均値(pooled mean of effect size)を計算した。

主な結果

レビュー結果は、21 件の研究論文ないし報告書にある 28 の調査コーホートによる 1,676 人の調査協力者からのデータをベースにしたものである。これらの調査は---多くがサンプル数は比較的少なかったが---ボランティアによる個別指導プログラムが生徒の言語学習能力と読解力のアウトカムにプラス影響を及ぼしうることを明らかにした。我々が最初に行ったのはボランティアによる個別指導が生徒の読解力のアウトカムの測定に及ぼす全体的な効果のチェックである。25 件の調査が何らかの種類の読解力の測定を評価していた。初等教育段階の生徒に対するボランティアによる個別指導プログラムの読解力に及ぼす平均効果量は 0.23 である。全体的結果とかなりかけ離れた外れ値を出していた 1 調査を除いた平均効果量は 0.30 となった。特定の学問分野のメタ分析においてもいくつかの顕著な結果が見つかった。ボランティアによる個別指導プログラムが顕著な違いをもたらしたところは、全体的読解力(Reading Global)(効果量= 0.26)、文字と単語(Letters and Words) (効果量= 0.41)、滑らかな発音力(オーラル・フルーエンシー/Oral Fluency) (効果量= 0.30)、及び作文力(Writing) (効果量 = 0.45)であった。ボランティアによる個別指導の読解力-理解力ならびに算数(Reading -Comprehension and Mathematics)に対するアウトカムはプラスであるが、あまり顕著なものではないことが分かった。

読解力の成績を調査の特性別に分析すると、指導者のタイプ、学年、あるいはプログラムの重視点によって、効果量についての有意差は生じないことが分かった。高度に構造化された個別指導プログラムは、あまり構造化されていないものに比べると、全体的読解力のアウトカムに対し、はっきりした有意差を示す効果があったが、その他のタイプのアウトカムでは差はなかった。学術雑誌に出版された調査と出版されなかった調査の間の効果量の差は統計的に有意ではなかった。出版バイアスに関するその他のテストも、レビューの対象となった調査がバイアスのないサンプルであることを示唆するものであった。

レビューワの結論

本レビューは、全国の学校や学校区で早速採用するよう推奨すべきボランティアによる個別指導

の具体的モデルがあることを示唆するものではない。より正確な言い方をすれば、本分析から我々が出せる結論は、これらの指導プログラムが低学年の児童の重要な読解力や言語の副次的能力(sub-skills)にプラスの影響を及ぼしうるということである。標準偏差の約3分の1---という結果は相当なものである。結果として、本分析結果は、ボランティアによる個別指導を低学年の児童の学力向上戦略の一つとして採用すべきか否かを決定しようとする立場にいる政策立案者や教育専門家にとって活用しうる重要なエビデンスの一端として役立つ。全国の教育専門家達が、国の説明責任制度(accountability system)にあって、1年ごとの発展目標を達成しようと努力し、毎年提示される学力水準目標に到達できない問題状況にある生徒に、手近にできる付加的な教育サービス提供を検討している現段階において、児童の読解力と言語能力を高める方法として、構造化され、読解力に重点を置いたボランティアによる個別指導プログラムは一考の価値がある。

1. 背景

多くの文化の中でも、最古の指導形態は子供に家庭教師(in-home tutors)又は個人的に雇用する指導者(private instructors)をあてがうものであった(Shanahan, 1998)。個別指導は世界中で良く知れている指導形態で、個別指導の教授法としての有効性は各種教育文献に幅広く記録されてきた(例えば、Cohen, Kulik, & Kulik, 1982 ; Fashola, 2001 ; Wasik & Slavin, 1993 参照)。

1970年代、米国の学校は、教員不足時代の限られたリソースの有効活用法の一つとして、仲間による指導(別名、生徒による生徒指導ないし同学年同士の(cross-age)指導)に対する依存度を高め始めた(Rekrut, 1994)。それに引き続く20年間も、成人のボランティアで構成された指導プログラムへの関心が高まるということが見られた。その理由としては、(1)「教育の卓越に関する国家委員会」(National Commission on Excellence in Education)が1983年に出した『危機に立つ国家』(A Nation at Risk)と題した報告書に以後の一般市民の教育の質に対する危惧の高まり；(2)1990年代の社会奉仕活動(community service)への関心の高まり、及び(3)専門の個別指導教員による、費用はかかるが効果的なプログラムによって有望な結果が得られたこと、等がある。米国学術研究会議(National Research Council)の推定によると、1987年までの時点で週平均4時間米国の公立学校でボランティア教員として働いた人の数が百万人を超している。同調査によると、米国の公立小学校の4分の3が、1校平均24人の割合でボランティア教員がいると答えた(Michael, 1990)。

その後、この種の大学との連携は、1997年クリントン大統領が発足させた全国的規模の個別指導の取組みとしての“アメリカ・リーズ・チャレンジ(America Reads Challenge)”によりさらに加速した。当時、全米の約800校に上る高等教育機関が学生達に小学校の生徒の個別指導教員として学校で働く実習枠(wor-study slot)で派遣する誓約をした(ホワイトハウス、1997)。1999年までに、約1,200校に上る高等教育機関が実習生を個別指導教員として公立学校に送り込む約束をしたのである。クリントン大統領の1999年度予算案にも、中退率が高く成績の悪い生徒が集中する学校の生徒に対して、大学からメンターを紹介するプログラムを制度化するために1.4億ドルが盛り込まれた。アメリカ・リーズ・チャレンジにより州の指導層の小学生向け個別指導プログラムに対する関心が深まり、今では多くの地元の個別指導の取組みが次第に支持を増やす傾向にある(米教育省1996；1997)。

過去数十年にわたっての個別指導プログラムに対する関心と支持の高まりとは裏腹に、正規ではない成人ボランティア教員を利用するプログラムはそれを支える研究基盤をとまってはこなかった。1970年代と1980年代の初旬、米国のほとんどの個別指導研究は仲間による個別指導や縦割り(cross-age)の個別指導の影響に焦点を当てていた(Shanahan, 1998)。1982年、Cohen, Kulik, and Kulikは米国教育研究ジャーナル(American Educational Research Journal)における初期の一連の個別指導研究に関する、その後よく知られよく引用されるようになったレビューを発表した。著者たちは、今まで何百に及ぶ個別指導に関する報告書が、教員や研究者から出され、科学的に正しい実験計画の評価に基づいて書かれたものもあれば、よりインフォーマルで主観的なものもあると指摘している。彼らによれば、1982年にまでの研究に対する主なる4つレビュー^{*1}は“比較的インフォーマルな体験談でありボックススコア技法(box score technique)”を用い、それぞれが仲間

^{*1} 4つのレビューとは：Devin-Sheehan, Feldman, & Allen, 1976；Ellson, 1976；Fitz-Gibbon, 1977；及びRosenshine & Furst, 1969を指す。

による個別指導は低学年の児童の学力を向上させるのに役立つという結論を出していた。

1980年代と1990年代、研究家は最も学力達成度の低い子供達の学力向上を目指した専門的介入に力点を置き始めた。その結果できあがった最も注目に値するプログラムは、読解力回復(Reading Recovery)プログラムと落ちこぼれ絶滅 (Success for All) プログラムである(Shanahan, 1998; Wasik & Slavin, 1993)。両プログラムとも、専門の指導員によるマンツーマンの個別指導を含んでおり、教育研究界の識者の多くが効果的と認めているものである(Wasik, 1998)。が、専門の指導員を雇用する場合、費用の関係で介入の対象となる子供の数が限定されるので、近年では、成人のボランティアもしくは補助的な専門家を個別指導員として活用するいくつかのプログラムがつくられた。これらのプログラムの一部は2001年の「落ちこぼれ防止法」(No Child Left Behind Act of 2001)が定める説明責任(accountability)措置への対応として作られたもので、教育行政の担当者に生徒の学力向上を目指した個別指導プログラムの利用を働きかけるものである。教育行政の担当者は、自分達の学校で個別指導プログラムの利用を推進する一方で、個別指導プログラムが現実的にどのような種類の効果を生み出すことができるのか確証を持たずに利用していただだけの可能性もある。個別指導プログラムの効果の理解に向けて、手始めにレビューのために既存の文献を検討した。

先行研究として、ボランティアによる個別指導プログラムの生徒の実績に及ぼす影響調査に関する4つの主なるレビューを特定した：Topping and Hill (1995), Wasik (1998), Shanahan (1998), 及び Elbaum et al. (2000)である。

1995年、Topping and Hill は『指導員・メンター役の学生たち』(Students as tutors and mentors) (Goodlad, 1995)と題する本に寄稿した章の中で、ボランティアによる個別指導に関する研究の興味深い背景を提示している。同章では、児童に対する個別指導員としての大学生の有効性に関する国内外の評価研究のレビューを行っている。

Wasik (1998) は、生徒の読解力向上を助けるボランティア指導員を利用する17のプログラムのレビューを行っている。ボランティア指導員が多くの児童の読解力向上に本当に役に立つことができるというエビデンスらしいものがある一方で、結果にはプログラム毎にかなりのばらつきがあった。さらに、レビューの対象となった17プログラムのうち、児童の達成度を対照群と比較しているものは2つだけで、それは、その結果をさらに不確実なものにしてしまっている。Wasikの調査結果は米国学術研究会議の1998年の報告書『児童の読解力不足の防止のために』(Preventing Reading Difficulties in Young Children)の結論と一致している。同報告書の結論とは、「ボランティア指導員が深刻な読解力の問題を抱える児童に効果的に対応できるということを確認しうるエビデンスは存在しない」である。とはいえ、これらの非系統的レビュー2つのうち一つは含まれておらず、その調査結果を改ざんしたといっても過言ではない。

対照的に、Shanahanのボランティアによる個別指導に関する研究についてのレビュー(1998)では、多くの条件的な制約があるにしても、これらのプログラムが生徒の学力向上に効果を持つことが分かったと言っているが、Shanahanは自分のレビューに用いた方法論についての詳しい内容にはほとんど触れておらず、上述のレビュー同様、いくつかの重要な調査結果を無視している可能性も十分ある。

ごく最近では、Elbaum et al. (2000) により、教育心理学ジャーナルに載せた読解力不足で問題を抱える小学生に対し読解力を向上させるマンツーマン個別指導プログラムの効果に焦点をあてた、メタ分析に関し報告がなされている。著者は、1975年から1998年の間に発表された42人の生徒をサンプルとする29の調査研究のレビューを行い、研修を受けた指導員であれば生徒の読解力向上に働きかけうることを明らかにした。このレビューは、役に立つことは確かであるが、唯一読解力だけを主題として採り上げ、個別指導の種類には熟練の専門教員によるものも含むあらゆるタイプのものを含めている。

ボランティアによる個別指導プログラムの有効性に対し、エビデンスが乏しく、往々にして矛盾があることは、我々に多くの疑問を残す結果となっている。その筆頭に挙げられるべき疑問は、政策立案者や学校経営者達が児童の学力向上に可能性のある方法としてボランティアによる個別指導を引き続き推進することを是とする良いエビデンスは果たして存在しているのか？参加者間の(年齢、性別等々の)差がプログラムの有効性に及ぼす影響は？そして個別指導プログラム間の(プログラムの構造、プログラムの主眼点、指導員のタイプ等の)違いが結果にどのような影響を及ぼすのか？といったことである。

2. 目標

ここでの系統的レビューの目標はボランティアによる個別指導プログラムに関する実験や経験に基づいた調査結果を収集・要約・統合し、政策立案者、教育者、両親、及びその他の関係者にこの種の介入が小学生児童の学力向上に有効なツールとなり得るものであるか否かを判断するための資料として役立ててもらおうことであった。

3. レビューの方法

レビュー対象調査の採用規準

調査の種類

無作為実地試験のみが本レビューの対象となった。処遇グループと対照群を主たるアウトカム変数の予備テストに当てはめて用いた準実験調査(quasi-experimental studies)は対象外とした。予備テスト-事後テスト調査、もしくはある処遇グループが別の処遇グループと比較される場合も対象外とした。1985以前に行われた調査も対象外とした。本レビューに対する限られたリソースのため、米国で実施されたプログラムのうち英語での調査のみが考慮対象となった。さらに、英語力が低い(LEP)生徒のニーズに対応するために特別に設けられたプログラムの調査も対象外とした。このような特別なプログラムはほとんどの小中学生向けの個別指導プログラムを代表するものとは言えないからである。

参加者の種類

我々のレビュー対象となったのは正規職員ではない成人の指導員が参加するプログラムの調査の

みである。これらの指導員は、調査文献の中ではほとんどの場合「ボランティア」と呼ばれていたが、(学部学生で、実習プログラムの一環として個別指導していた)指導員に少額の謝金を支払っていたプログラムも対象とした。この領域に属するプログラムに、自分達の子どもを生徒として指導する保護者の指導員を養成するプログラムがいくつかあったので、これらも本レビューの対象に含め、下位グループ分析として別途検討するためコード化した。個別指導の対象となる生徒については、K-8 学年生(小・中学生)を指導対象とするプログラムの調査のみを本レビューの対象とした。その理由は、小中学生がボランティアによる個別指導プログラムの最も一般的な対象であったことと、このプログラムが高校生を対象とするプログラムとは根本的に異なるものであるからである。

介入の仕方

介入の仕方の特徴は、少なくとも 1 ヶ月間継続の学力向上を主眼とする定期的個別指導講座である。この期間制限を設けた理由は、数日で終わってしまうプログラムは、持続的影響を受ける長期継続プログラムと質的に異なると考えたからである。介入の主目的については、学力向上以外の要素(例えば、行動変容)も含めることにしたが、評価をするためには少なくとも一つの学業成績のアウトカム指標に焦点を当てざるを得なかった。

成績のアウトカム指標の種類

本レビューの当初に意図していたことは、遠隔の (distal) アウトカム(例えば、学校での成績や標準化された達成度テストの得点を上げる等のプログラムが実際に影響を与えたいと意図しているもの)と近接した (proximal) アウトカム(例えば、生徒の出席率のような、個別指導の影響で変化し、将来の実績向上につながる可能性のある中間的指標)を含む生徒の達成度に関連するあらゆる成績の指標を考慮対象とすることであった。しかし、学校の成績や出席率の分析をしているレビューは非常に少なく、ほとんどの研究は計算力と読解力の標準化された評価(standardized assessments)、又は読み書き能力の“信頼性ある”(authentic) 指標に照準を合わせていた。そのため、我々のレビューもこれらの成績のアウトカムを中心にして行うこととなった。

対象となった調査の多く、特に読解力を中心にしたものは、複数のアウトカムを調査対象としており、これらのアウトカムを我々のアウトカム指標の分類に従って区分けする作業は我々レビューワの仕事となった。我々は本レビューの対象論文に出てくる各種コンセプトに基づき、アウトカムを 6 つの大きなカテゴリーに分類した。次に掲げるのは、その 6 つのカテゴリーと、各カテゴリーに合ったアウトカム指標である。(セクション 12 の表 11 は、各カテゴリーのアウトカム指標とその大きなカテゴリーに入る各調査から採りだした個々のアウトカムを示している。)

全体的読解力(Reading Global)。このカテゴリーには、数多い標準化された読解力達成度テストの結果を含めた。このカテゴリーの達成度テストには、ゲーツ・マクジニシー(Gates MacGinitie)読解力テスト、ワイド・レンジ達成度テスト(Wide Range Achievement Test)読解力セクション、基礎的スキル総合テスト(Comprehensive Test of Basic Skills)読解力セクション、スタンフォード達成度テストの読解力バッテリー(一連の部分テスト(=サブテスト)より成る関連テスト集)等が含まれる。

文字・単語読解力(Reading Letters and Words)。多くの読解力を中心とする調査は、我々レビュー

ワがここで“文字・単語”カテゴリーに分類する読解力サブスキル(下位能力)関連の複数のアウトカムを調査対象としていた。このカテゴリーに入る指標のタイプは単語の解読(decoding)や言葉についての知識に焦点を合わせているものである。この分類法の基底には、生徒が上手に読めるようになるためには特定の能力が必要で、その能力は単語を読み、その意味を理解できることに関する能力との論理がある。

レビューワがこのカテゴリーに含めたアウトカム指標は、一般的な読解力の標準化された評価法に属するサブテスト全般である(読解力のサブテスト付きの標準化された評価法としては、ゲーツ・マクジニシー読解力テスト、ワイド・レンジ達成度テスト読解力セクション、基礎的スキル総合テスト、ウッドコック・ジョンソン教育心理学バッテリー、単語読解能力テスト等がある。) 読解力のアウトカム指標の例としては、単語の識別(Word Identification)、ワード・アタック(Word Attack)、文字の識別(Letter Identification)、ディベルズ発音力評価(DIBELS fluency assessments)、単語認識(Word Recognition)、語彙テスト(Vocabulary tests)等がある。このカテゴリーには又、子音の発音、短母音、連字やコンビネーション、視覚語(sight words=訳者註：“and, but, there, blue”等々の丸暗記語)、及び非語解読(nonword decoding)をテーマにしたサブテストも含まれる。これらの指標には規格化されていないものもある。例えば、モリス他の調査 1990(Morris et al. 1990 study)では標準化された評価法と何の関連性もない単語認識アウトカム(word recognition outcome)を活用している。

カテゴリーの分類の仕方は数多くあるが、目標とするところはアウトカム指標に類似のタイプを含んだ、カテゴリーの正当な数を整理することであった。このカテゴリーでは、低学年児童がすらすらと読めるようになるために特に必要とされるサブスキルが中心となる。

読解力.このカテゴリーには、標準化された読解力達成度テストの理解度サブテストの結果が含まれる。本レビューの対象調査に使われている理解度サブテストは、ゲーツ・マクジニシー(Gates MacGinitie)読解力テスト、ワイド・レンジ達成度テスト、基礎的スキル総合テスト、スタンフォード達成度テスト、及びウッドコック・ジョンソン教育心理学バッテリーからのものである。

読解力オーラル・フルーエンシー(滑らかな発音力).読解力を中心とする調査の多くは生徒の文章の一節を早く正確に音読できる能力を測るアウトカム指標を使っていた。このようなアウトカムは概して生徒にある文章の一節を読ませ正確に読まれた単語数に基づいて生徒を評価している。このカテゴリーに含まれるアウトカム指標としては以下のものがある：履修課程ベースのオーラル・フルーエンシー(curriculum-based oral fluency)、基礎的引用文集(basal passages)、読解力レベル観察調査(observational survey – reading level)、分析的読解力目録(流暢さ)(analytic reading inventory (fluency))、等々。

作文力.本レビューの対象調査のうち、作文力指標というカテゴリーのアウトカム指標を使っていたのは 6 調査だけであった。このカテゴリーに入るアウトカム指標としては以下のものがある：スペリング、観察調査－作文力(observational survey – writing)、観察調査－口述筆記(observational survey – dictation)、ワイド・レンジ達成度テストのスペリング・サブテスト、特定作文サンプルにおける綴りに間違いなく正確に書かれた語数、カリキュラムに基づくスペリング指標

(curriculum-based spelling measure)。

全体的算数力.本レビューの対象調査のうち、算数力指標というカテゴリーのアウトカム指標を使っていたのは 5 調査だけであった。このカテゴリーに入る算数力達成度指標としては以下のものがある：研究者が開発したかけ算テスト、スタンフォード達成度テストの算数サブテスト、及びオルレアン・ハンナ代数予知テスト(Orleans-Hanna Algebra Prognosis Test)。

調査の特定に利用した検索方法

本系統的レビューの対象となるボランティアによる個別指導プログラムに関する調査タイトルの絞り込みは、いくつかの方法を使って行われた。レビューワは先ず、C2-SPECTR, EbscoHost Research Database の検索を行った。その時に使用したデータベースは以下の通り：Academic Search Premier；Primary Search；Professional Development Collection；Middle Search Plus；Psychology and Behavioral Sciences Collection；PsycINFO；ERIC の Sociological Collection、及び Proquest Digital Dissertations。

当初の検索は、以下のようなキーワードを様々に組合せたり、切り離したりして行われた：“ボランティア又はメンター(mentor)又は個別指導員(tutor)*又は個別指導プログラム(tutorial programs),”及び“初等又は初期教育又は中学生又は初期介入,”及び“コントロール又はランダム又は実験又は評価又は非仲間プログラム”。コンピュータ・データベースで未更新のままになっている論文を探すために、我々は、Education Next、Education Policy Analysis Archives (EPAA)、Educational Evaluation and Policy Analysis (EEPA)、読解力調査季刊誌(Reading Research Quarterly (RRQ))、及び教育研究レビュー(Review of Educational Research (RER))等の我々の調査に最も関連性のあるいくつかの一流誌の、2003-2005 年の間に出版された分の目次のオンライン検索を行った。レビューの対象は、米国で英語を母国語として用いる人を対象に実施された調査に限定した。教育政策との関連性を保つために、検索は米国内で実施された調査に関する論文のみを対象とすることにした。その理由は、米国外での教育制度の中で実施された介入が米国の教育政策環境にうまく適合するとは考えられなかったからである。

上記検索結果によりできあがった論文リストに、上述のボランティアによる個別指導に関する 4 つの広く引用されているレビューが参照しているその他の調査研究、即ち、Elbaum et al., 2000；Wasik, 1998；Shanahan, 1998；及び Topping and Hill, 1995 が付け加えられた。レビューワは又、検索作業の精緻化を図るために、情報スペシャリスト、司書、読解力スペシャリスト(reading specialist)、及びキャンベル・コラボレーションの情報検索政策概要(Campbell Collaboration's Information Retrieval Policy Brief) (Rothstein, Turner, & Lavenberg, 2004)等の情報筋の助言をも求めた。調査の主な収集先となったのはアーカンソー大学図書館システム(University of Arkansas Library System)、図書館相互貸借(Interlibrary Loan)、大学所蔵マイクロフィルム(University Microfilms)、及び上述のデータベースである。すべての調査タイトル(題名)と対象指定理由はレビューの正確性と一貫性を維持するために表計算ソフト“エクセル”を使って記録され、管理された。可能な限り、すべての論文の PDF ファイルが中央ネットワーク・フォルダーに保存された。これらの論文のハードコピーとプリントアウトの形でしか入手できなかった論文(print-only articles)もファイルに保存された。

この過程で作成された調査タイトルのリストは、少なくとも二人のレビューによる調査内容の概要チェックにより、絞り込まれた。概要の収集とチェックが済むと、両レビュー共選択した全調査の全文に目を通した。

この第1回全文チェックに合格した調査は完全コード化段階に回された。これらの調査の(少なくとも二人のレビューによる)完全コード化が済んで、すべての採用規準に合格して残った最終組の調査は、引き続き分析され、その結果は総合された。コード化過程で二人のレビューが異なる結論に至った場合は、その論文を残すか、落とすかはこの二人のコーダー(コード化作業員)が折り合せて決定した。この折り合い過程の中身は、コーダー同士が該当の論文を残すか・落とすかの是非を、合意に至るまで論じ合うことであった。合意に至らなかった場合のデフォルト値は“該当の論文を残す”であった。最終的には、主任レビューがすべての意見の不一致を收拾した。

調査のデザインに関しては、無作為のものが因果関係推定(causal inferences)を行う基盤として最も強力なデザインであると広く受け入れられている。入手できた文献の初回探査で数多くの無作為実地試験が見つかった。レビューの対象選びは、入手可能な最も信頼性のある証拠物件の利用を最優先すべしと言う前提に基づいて行われた。無作為実地試験が僅かしか見つかっていなかったとしたら、高質のな半実験的デザインのものを対象に選ぶことになったであろう。しかし、無作為実地試験に基づく20件を上回る調査や報告の中に約30に上る調査コーホートを特定できたので、我々は半実験的デザインのものは我々のメタ分析から除外することに決めた。

試験の選択

次の表1が示すように、検索データベースで、検索条件に基づいて最初に特定できた調査は合計1,437件に達した。但し、中には多重検索によって重複して出てきた調査も混じっていたので、チェックの対象となった調査概要の最終件数は969となった。

表 1：最初の検索条件と検索結果

検索先	検索条件	調査概要の件数
C-2 SPECTR	(全索引付きフィールド) tutor(個別指導員) OR (全無索引フィールド) tutor AND 1985-2005 AND volunteer(ボランティア) OR mentor(メンター) OR adult(大人・成人)	12
EBSCOhost Research Database (内訳 Academic Search Premier ; Primary Search ; Professional Development Collection ; Middle Search Plus ; Psychology and Behavioral Sciences Collection ; PsycINFO ; Sociological Collection)	(テーマ用語) “Volunteer or mentor or tutor* or tutorial programs(個別指導プログラム)” and “elementary or primary education(初等教育) or middle school students(中学生) or junior high school students(中学生) or early intervention(早期介入)” AND (デフォルト値フィールド) control(制御) or random(無作為) or experiment(実験) or evaluation(評価) or program not peer(非同僚プログラム)”	268
ERIC	(SU Descriptors(記述子) All(全)) “Tutor(個別指導員)* not peer(非同僚) not college(非大学) not faculty(非教員)” and “primary education(初等教育) or elementary(初歩) or elementary secondary(初等中等) or junior high or middle school students(中学生) or early intervention(早期介入)” AND “control(対照) or random(無作為) or experiment(実験) or program(プログラム) or evaluation(評価)”	543

Proquest Digital Dissertations (プロクエスト・デジタル論文)	(キーワード) “tutor? or tutorial program(個別指導プログラム)” AND (AB) “primary education or elementary or middle school students or early intervention” AND (AB) “control or random or experiment or program or evaluation) AND (date) “>=1985” AND (date) “<=2005”	76
Review of Reviews (各種レビュー のレビュー)	1985年以降出版された論文。ここでレビューされる4つの論文と は: Elbaum, Vaughn, Hughes, & Moody (2000); Shanahan (1998); Topping & Hill (1995); and Wasik, (1998).	167
Hand Searches (手検索)	期間 (2003-2005) - Education Next(来るべき教育), Education Policy Analysis Archives (EPAA=教育政策分析アーカイブズ), Educational Evaluation and Policy Analysis (EEPA=教育の評価と政策分析), Reading Research Quarterly (RRQ=読解力研究季刊誌), and Review of Educational Research (RER=教育研究レビュー)	371
	回収した概要の総数(重複分込み)	1,437
	回収した概要の総数(重複分削除後)	969

調査タイトルのリストは、コンピュータ・データベース検索、手作業による検索(hand searches)及び各種レビューのレビュー(review of reviews)によって作成された。各タイトルの概要が収集され、二人のコーダーが概要を読み該当の調査が採用規準を満たしているか否かを決定した。これらレビューは、概要の絞り込みにあたり、以下の指針をベースにし、該当するものを不採用とした：

- ・ 当論文はメタ分析の報告を載せている。
- ・ 当調査は実験的デザインを使用していない(例えば、事例研究(case study)、体験談(narrative)、長さ約 1-2 ページの簡潔な分析/レポート、あるいは質的調査(qualitative study))。
- ・ 当調査は 1985 年以前に実施された。
- ・ 生徒は米国外に居住していた、あるいは英語を母国語としていなかった(即ち、ESL(第 2 言語としての英語)、LEP(限定的英語熟練))
- ・ 個別指導員は仲間もしくはプロの教員であった(即ち、ボランティアによるの成人指導員ではなかった)。
- ・ 生徒の学年は K-8(幼稚園から 8 年生)の範囲からはみ出していた。
- ・ 介入の主目的に少なくとも一つの学力成績の向上が含まれていなかった。
- ・ 介入期間が 4 週間もしくは 1 ヶ月を下回っていた。
- ・ 特殊な生徒を対象としていた(即ち、聴覚障害者、視覚障害者等)

検索結果から上記不採用規準に明確に合致している概要を持つ調査が取り除かれた。概要だけでは情報不足で判断できなかった調査は、第 2 ラウンドの分析に回された。

論文の過半数はこの第 1 ラウンドの選考で没になった。その理由は：実験的デザインを使っていなかった；指導員がボランティアではなかった；生徒が英語を母国語としていなかった、もしくは米国外に住んでいた；あるいは生徒が K-8 学年に属していなかった。この篩い分け段階でかなりの数の調査が絞り込まれた；969 調査中、最初の論文概要審査に合格したのは 233 件のみ。

二人のレビューは残った 233 論文を初回全文チェックにかけた。全文チェックで最初の評価対象となったのは論文の“紹介”と“調査の方法”欄であった。この段階の目的は、残った論文の完全コード化に取りかかる前に各論文についての理解を深めることであった。レビューは先ず論文が何らかの調査結果を報告しているのものであって、単なる特定のプログラムの内容説明ではないことを確認した。次に、“調査の方法”欄を調べ、生徒の処遇グループと対照群への振り分けが無作為に行われたことを確認した。最後に、レビューはここでも上記の全般的な不採用規準に基づく論文の絞り込みをおこなった。これは、概要(abstract)欄より紹介 (introduction) 欄の方がより優れた情報を与えてくれることが屢々あったせいである。この段階でさらに絞り込みがかなり進み、完全コード化段階に回されたのは 56 論文のみとなった。

最後に、この最終完全コード化段階で残った 56 論文の半数以上がさらに切り落とされる結果となった。結局、我々のメタ分析の対象として残ったのは 21 論文のみ。但し、残った論文中数件は別個のコーディングを必要とする複数のコーホートを持っていたので、これらの論文から得られた調査コーホート数は 28 となった。例えば、Cobb 2000 論文は 3 つの各学年の結果を別々に報告しているだけで、プールされた(共有)データは一切提供していない、従って、我々はこの調査の結果を 3 つの別個の Cobb 2000 調査コーホートとして我々のレビュー対象に組み入れた。

最終完全コード化段階で合計 35 論文が落とされた。落とされた 35 論文中 30 件の不採用理由は：

介入期間；真の無作為実地試験調査のデザインとは言えない；プログラムが実施された場所は米国外である；関連性のある学力成績がない；個別指導が対面式でない；及びボランティアというよりはむしろプロの指導員を使っている、であった。残りのメタ分析の対象から外された5論文の不採用理由は、これらの調査に算出されている統計値の質の不安があったせいである。これらの5調査が報告している統計は、我々のメタ分析の対象にするには不十分であった。

我々は当初準実験的デザインも含める予定だった。しかし、(因果関係推論の基盤とするのに最も強力なデザインである)無作為デザインを使っている調査が21件特定できたので、準実験的デザインをメタ分析から除外することにした。

調査方法の質の評価

各調査(とその報告の仕方)の質の評価の評価対象となった特性としては：1) 調査の透明性(transparency)；即ち、レビューが当該無作為振り分けの仕方を説明している報告内容の透明度；2) 無作為振り分けデザインの完全性(integrity)ならびにその研究計画に対する違反(violations)へのレビューによる対応の有無；3) 当初の無作為振り分けサンプルからの生徒もしくは指導員いずれかの高レベルな自然減少の発生；4) 処遇の忠実性(treatment fidelity)に関するかなりの数に上る問題の発生、等である。各調査の質の評価はコード化のレビュー会で4人のレビュー全員で行った。

メタ分析の対象となった21調査のうち、生徒の処遇グループと対照群への振り分けに問題と言える問題が発生したのも、自然減少を疑わせる兆候を見せたものも無かった。忠実性(fidelity)に関する疑問がレビュー対象となった調査の数件について提起された。これらの疑問はレビューの対象となった各調査の詳細を要約した表4に載せた。しかし、レビューはそれらの調査のいずれをも対象から外すまでに至らなかった。

調査の質につき最後に採り上げられた問題は調査の統計値の報告の仕方に関連するものであった。除外された6調査は、少なくとも部分的に、レビューが標準化された効果量の平均差(standardized mean difference effect sizes)を計算できるに足る適切な記述統計量(descriptive statistics)を記すことを怠っていた。うち、特に2調査(Compton, 1992; Meier & Invernizzi, 2001)は必要な記述統計計算を可能にする推計統計を記していたが、その記載値はつじつまの合わないものであった。Compton 論文の合計サンプル・サイズは483に上っていたが、同論文に記されている標準偏差は報告されている標準誤差あるいはt値に合致せず、標準偏差の約2倍に匹敵する効果量を記していた。Meier and Invernizzi 論文には、平均値が出ていたが標準偏差がなく、報告されているF値は同調査のサンプル数に矛盾する自由度(df)を持つものであった。

データの管理と抽出

当初レビューした234の論文中、109論文はオンライン・ソースから、66論文はアーカンソー大学図書館からマイクロフィッシュ、マイクロフィルムあるいは製本された雑誌の形態で、(学位論文を含む)58論文は図書館相互貸し出しにより取り寄せ、1論文(学位論文)は貸し出し中で当図書館からすべての可能性のある行き先を調べたが見つからなかったため入手できなかった

た。各論文の電子コピーあるいは印刷コピーを入手後、我々はある指定されたコード化フォームを使って各論文から情報を抽出した。レビューチームの全員がコード化フォームの作成に加わった。調査のコード化、データ管理、及びメタ分析のすべては表計算ソフト“エクセル”を使って行われた。

データの統合

我々は、各アウトカム指標に標準化された平均差効果量統計値のヘッジズの不偏推定値 g (特定の結果変数に対する処遇グループと対照群の平均値間の差を後指標用にプールされた標準偏差で割った値) を用いた。平均値や標準偏差値の記述がないときは、我々は効果量の推定値をウィルソンとリプシー(2001)が勧める手段を使って推定した。事前・事後指標の双方を入手できた際は、我々は予備テストのグループ差を差し引いた。というのも、無作為に振り分けられたグループであったため、ほとんどの場合グループ差は無いに等しいものであったからである。

調査が事後テストの調整平均値か後テストの平均値しか報告していない時は、我々は処遇グループ対対照群間差^{*2}を計算した。いずれの場合も、我々は後テストの得点のプールされた標準偏差を効果量 d を算出する際の分母として使った。人口効果量(population effect size)の不偏推定値(unbiased estimates)は、Hedges and Olkin (1985)の近似値(approximation)で割り算して算出した： $1 - 3/(4N - 9)$ 。

一部の調査はプログラムの効果を評価するために様々なアウトカム指標を使っていた。算数の成績は言語学習の成績とは質的に異なるので、我々は各調査の“効果”もすべての入手可能調査の“全体的効果”の計算も行わなかった。しかし、読解力の指標に対する全体的効果の計算は行った。我々は、それをするために、何らかの種類の読解力成績が評価されていた 25 調査コーホートそれぞれの全体的読解力効果量を計算した。次に、特定の介入がある一つの分野でより大きい効果を持つか否かを決定するために、標準化された全体的読解力、標準化された文字・単語力、標準化された読解力、音読力(oral reading fluency)と作文力の指標、標準化された算数の能力等を含む重要成績分野のメタ分析を別個に行った。特定の重要成績を複数の方法を使って測定している調査については、我々は複数の指標に基づく効果量の平均を出して、各重要成績の分析における調査 1 件あたりの寄与度は 1 データポイントに限定し、メタ分析に特定の調査が不当に“重用される”ことのないようにしなければならないとする原則を守った。

等質性分析

等質性分析テストをすると、効果量のばらつき(variations)がサンプリングの仕方が間違っただけか、それとも別の要因によるものであるかが分かる。固定効果モデル(fixed effects model)

^{*2} 本検討の分析対象になっている 28 件の調査コーホートのうち 12 件は事前テスト格差(獲得点数と同等)に対して調整した事後テスト得点を使って分析し、3 件は(ANCOVA を使って)事前テスト格差に統計的に調整した後テスト得点を使って分析し、12 件は前テストの点数がなかったため後テストの得点のみに基づいて分析し、1 件は後テストの得点のみと調整したテスト得点のミックス(混合)に基づいて分析した。一つの共通点は効果量の計算に用いられた分母である。いずれの計算においても、効果量の統計値の分母はプールされた後テストの標準偏差であった。

か無作為効果モデルかのいずれを使用するかは等質性分析に基づいて行われる。全体的効果と 6 つの重要成績の分析が明らかにしたのは、等質性の帰無仮説を退けるに十分とは言えない大きさの Q 統計値であった。つまり、効果量間の可変性はサンプリングの仕方の間違いに基づく期待値を超えてはいなかった(Lipsey & Wilson 2001)。従って、我々の研究ではデータの総合に固定効果モデルを使用した。

感度分析

我々は、包括的メタ分析ソフト（Comprehensive Meta Analysis ® software）を使って我々の得た主な結果のメタ分析対象調査に対する感度テストを行った。“1 調査ずつを除いていく”分析（“one study removed” analysis）を行うと、分析の対象から 1 調査ずつ順次除いていった残りの全調査の平均標準化平均差(average standardized mean difference)が得られる。最終的にこの感度分析で明らかになったことは、非常にサンプル数の多い 1 調査がメタ分析の結果に不釣り合いな影響を及ぼしているということであった。従って、本レビューに記す読解力成績に基づく結果はサンプルから Ritter-00 調査を取り除いた上で出されたものである。（感度分析で、Ritter-00 調査の全体的算数力の成績に対する影響は不釣り合いといえる程度のもではなかったため、算数の成績のメタ分析用としては、Ritter-00 調査はサンプルに残された。）

下位グループ分析

1) 指導員の種類、2) 生徒の年齢、3) 高度に構造化されたプログラム対構造化されていないプログラム、4) 出版元、等の各種プログラム特性に基づく微分平均効果量(differential mean effect sizes)を算定するために、下位グループ分析を行った。

つまり、我々が調べたことは、出版物に載せられる形式で公表された調査は、そうでないものに比べ、どちらかといえば、プログラムのプラス効果を示している可能性が大であると言えるかどうかであった。これらの下位グループの分析は、作文力と計算力の成績の分野では、それぞれ 6 件と 5 件しか調査がなかったため、本レビューの対象となっている 4 つの読解力成績についてのみ行われた。

出版バイアス

出版バイアスの測定は、trim and fill 法という、漏斗プロット(Funnel Plot)の相違点の外観検査によって行われた。さらに、我々は、出版物の種類をモデレーター(調節)変数(moderator variable)として出版バイアスの可能性を調べた。

調査データの記載不備

我々の調査のために残しておいた無作為実地試験が効果量の計算に必要な情報を十分に提供してくれた。

下位グループとモデレーターの事後分析(PostHoc Analysis)

メタ分析の一環として、我々は、調査そのものの中にある特性による効果量の差というものが存在するか否かを決定するために、いくつかの下位グループとモデレーターの分析を行った。この

モデレーター分析に含められたものとしては：1) 指導員の種類(保護者、大学、地域社会)；2) 学年、1年生あるいは他学年；3) プログラムの構造化度合い；4) 出版物の種類、等がある。

表 2： メタ分析の対象となる 28 調査コーホートのプログラム特性

	指導員 保護者/大学/地域社会	学年 1 含む?	読解力 が中心?	高構造化 授業 ?	雑誌掲載 論文?
Allor-04.1	0 / 1 / 0	1	1	1	1
Allor-04.2	0 / 1 / 0	1	1	1	1
Baker-00	0 / 0 / 1	1	1	0	1
Cobb-01.1	0 / 1 / 0	1	1	0	1
Cobb-01.2	0 / 1 / 0	0	1	0	1
Cobb-01.3	0 / 1 / 0	0	1	0	1
Cook-01.1	0 / 1 / 0	1	1	0	0
Cook-01.2	0 / 1 / 0	0	1	0	0
Cook-01.3	0 / 1 / 0	0	1	0	0
Erion-94	1 / 0 / 0	0	1	0	0
Mahoney-86	1 / 0 / 0	0	0	1	0
Mayfield-00	0 / 1 / 0	1	1	1	0
McKinney-95	0 / 1 / 0	1	0	0	0
Mehran-88	1 / 0 / 0	1	1	1	1
Miller-94	1 / 0 / 0	0	1	0	0
Morris-90.1	0 / 0 / 1	0	1	1	1
Morris-90.2	0 / 0 / 1	0	1	1	1
Nielson-92	0 / 0 / 1	0	1	1	0
Parham-94.1	0 / 0 / 1	0	0	1	0
Parham-94.2	0 / 0 / 1	0	0	0	0
Powell-Smith-00	1 / 0 / 0	0	1	1	1
Pullen-04	0 / 1 / 0	1	1	1	1
Rimm-Kaufman-99	0 / 0 / 1	1	1	1	1
Ritter-00	0 / 1 / 0	0	0	0	0
Vadasy-97a	0 / 0 / 1	1	1	1	1
Vadasy-97b	0 / 0 / 1	1	1	1	1
Vadasy-00	0 / 0 / 1	1	1	1	1
Weiss-89	0 / 0 / 1	1	1	0	0
合計	5 / 12 / 11	14	23	15	15

4. 結果

対象に選ばれた調査の説明

検索の結果得られたものは、21 件の論文、レポート、又は学位論文；これら 21 レポートの中で特定された“調査コーホート”あるいは“調査”の数は、一部レポートが複数のコーホートを別々に分析していた関係で、28。これら 28 調査の詳しい説明はセクション 9 の表 4 にあり。ここに説明する証拠基盤がよりどころとしているのは、1,676 名の調査参加者サンプル数、うち 873 名は個別指導の処遇グループ、803 名はその対照群に所属していた。

アウトカム指標。 本レビューの対象となったボランティアによる個別指導プログラムは、プログラムの有効性を評価するために様々なアウトカム指標を用いている。各対象調査の指標を見比べた上で、レビューはメタ分析に 7 つのカテゴリーのアウトカム指標を含めることにした。最初のカテゴリーは全体的読解力で、その証拠基盤を形成するのは、ある種の読解力の指標の評価を行っている 25 調査(サンプルの合計数 1,462 名の生徒)である。しかし、感度分析で 1 つの異常調査が見つかったので、我々はこれを取り除いた。従って、分析用サンプルに含まれる調査数は 24 件、合計生徒数は 1,077 人となった。プログラムの全体的効果を調べた後、我々は下記 6 つの

具体的学力分野(詳細については方法欄参照)の分析に取りかかった：

- (1) 全体的読解力：14 調査に基づいたエビデンス、合計サンプル数 819 人。
- (2) 文字・単語読解力：15 調査に基づいたエビデンス、合計サンプル数 798 人。
- (3) 読解力：8 調査に基づいたエビデンス、合計サンプル数 546 人。
- (4) 読解力オーラル・フルーエンシー：12 調査に基づいたエビデンス、合計サンプル数 635 人。
- (5) 作文力：6 調査に基づいたエビデンス、合計サンプル数 228 人。
- (6) 全体的算数力：5 調査に基づいたエビデンス、合計サンプル数 643 人。

上記調査数とサンプル数には重複しているところがある。その理由は、例えば多くの全体的読解力の成績を対象にしている調査が、読解力オーラル・フルーエンシー分野での結果も評価しているからである。

指導員の種類。 ボランティアによる個別指導プログラムが通常指導員を募る際の募集先は様々である。が、プログラムの中に若干数の特異なタイプのものであり、本レビューにおいては、これらを切り離し、別個に分析する。一部プログラムは保護者に指導員としての訓練を施し、彼等自身の子供達の個別指導に当たらせているのである。このプログラムは大学生を指導員として養成し彼らより若い生徒の世話をさせているプログラムとは異なる；往々にしてこれらの大学生は、America Reads プログラムの所属メンバーもしくは教職志願生である。最後に、本レビューの対象になっている残りのプログラムは、高学年の高校生から高齢者に亘る幅広い年齢層の地域社会ボランティアを利用してしている。この種の指導員の組合せを利用してしているプログラムは地域社会ボランティア・カテゴリーに入れた。我々の 28 件の調査“コーホート”のサンプル中、5 件は主として保護者を利用してしているプログラム(調査サンプル数=338)から、12 件は主として大学生指導員を利用してしているプログラム(調査サンプル数=899)から、残る 11 件は年齢がまちまちの地域社会ボランティアを利用してしているプログラム(調査サンプル数=439)から来たものであった。

生徒の年齢/学年レベル。 ボランティアによる個別指導プログラムの効果は生徒の年齢によって大きくなったり小さくなったりすると考えられる。従って、我々はレビューの対象とした個別指導プログラムを最も若い生徒(学年 1)向けのものとしてそれより高学年(学年 2 以上)向けのものとして分割した。我々の 28 件の調査“コーホート”のサンプル中、14 件は 1 年生(調査サンプル数 = 770)、残りの 14 件は 2 年生以上(調査サンプル数 = 906)を対象としていた。

読解力重視プログラム。 ほとんどのプログラムは読解力を含めることを重視していた：このようなプログラムは、読解力を特に重視しないプログラムよりは読解力の得点により大きな効果を発揮する筈であると常識的に考えられる。従って、我々は我々のレビュー対象となった個別指導プログラムを読解力重視型のものとしてより一般的な学力重視型のものとして分割した。我々の 28 件の調査“コーホート”のサンプル中、23 件は読解力重視型(調査サンプル数 = 1,033)、5 件はその他(調査サンプル数= 643)であった。我々の当初の計画では、プログラムの重視対象をモダレーター変数として利用する予定をしていた。ところが、読解力の成績を載せている調査が 1 件だけあったが、それは読解力重視型プログラムでなかった(McKinney-95)ので、下位グループ分析は行わなかった。

プログラムの構造。 調査は、指導員に対する命令や指図の多寡で高構造化型と低構造化型に分

類された。プログラムが指導員に具体的な授業内容と調べておくべき資料を提供している場合は、そのプログラムは高構造化型に分類された。生徒が自由に読み物を選択できるプログラムであっても、個別指導講座で各読書活動に費やす時間を指定している場合は、このプログラムも高構造化型に分類された。その他の、意図的に非命令調で指導員研修をほとんど提供しないプログラムや指導員と生徒がただ単に一緒に読むだけのプログラムは低構造化型に分類された。

例えば、ハワード通り個別指導プログラム (the Howard Street Tutoring Program) は、Morris, Shaw and Perney (2000)の説明に基づき高構造化型に分類された：“その 3:00-4:00 p.m. の個別指導時間は綿密に計画され、作業で詰まっております、中断は極めて少なく、1 時間のレッスンは通常次のような形式で進行する...” (pp. 136-137)。別の事例では、Vadasy, Jenkins, Antil, and Wayne (1997a)が健全パートナーズ・プログラム (the Sound Partners program) の初期バージョンをこう説明している、“その介入は 30 分間の放課後の補習レッスン 100 回が一つのセットになっており、個別指導員が 1 年生を対象に発音の仕方と初歩的な文章の読み方を教える内容のものであった” (p. 31)。

上記にひきかえ、読書家づくりを今日から始めようプログラム(the Start Making a Reader Today (SMART) program)は明らかに非構造化型であった。“指導員に与えられたものは授業に利用できる具体的なテクニックではなく、漠然とした枠組であった” (Baker, Gersten, & Keating, 2000, p. 497)。彼等は、指導員に専門的な教育能力を要求しなければ指導員の募集が楽になるとの提言も行っている。生徒が自分達の読み物を自分達で選択することができるペア読解法(paired reading)や反復読解(repeated reading)法も非構造化型に分類される (例えば、Miller, 1998 ; Weiss, 1989)。

我々の 28 件の調査“コーホート”のサンプル中、15 件は高構造化型に属するプログラムから来ており(調査サンプル数 = 919)、13 件はその逆であった(調査サンプル数 = 757)。

出版元。 系統的レビューとなると、“出版バイアス(publication bias)”又は“お蔵入りバイアス(file-drawer bias)”についての不安が現実となる。これらの専門用語が指しているのは、何の効果も示さない調査は、そうでない調査に比べ、他の点では同じであっても、提出された場合、受け入れられる可能性が少なくなるのではないかという考え方である。従って、雑誌に出版されるような調査は、出版されない形で頒布されるレポート、会議用資料、学生の学位論文等に比べプログラムのプラス効果を示すものになる可能性が大きいと考えられる。従って、この“バイアス(偏向)”のテストを行うために、我々は個別指導プログラムの調査サンプルから雑誌に出版された調査を抜き出した。我々の 28 件の調査“コーホート”のサンプル中、15 件は査読の対象となった雑誌から採ったもので(調査サンプル数 = 772)；残る 13 件の調査コーホートは主として博士論文から来ていた(調査サンプル数= 904)。

全アウトカムの総合効果量

我々は先ず生徒の読解力のアウトカム指標に対するボランティアによる個別指導の全体的効果を調べることから始めた。25 調査が何らかの種類の読解力指標の推定を行っていた。これら 25 調査中 8 調査の全体的読解力点数はある単独の読解力指標に基づくものであった；我々は、複数の読解力指標を使っている残りの 17 調査については、複数の読解力成績に基づく平均効果量を算出した。我々の分析によると、本レビューの対象となっている種類のボランティアによる個別指導

介入は、参加している生徒の言語能力に対し、かなりのプラス効果を持っていることを示している。図 1 は、ボランティアによる個別指導プログラムが小学生の読解力の成績に与える平均効果が、固定効果モデルを利用した場合 0.23 となることを示している。

次に、我々は感度分析を行って、このボランティアによる個別指導の読解力に及ぼす平均効果がある単独の調査結果の不釣り合いな影響を受けていないかどうかを調べた。その結果、Ritter-00 調査の持つサンプル数が全サンプル数の約 25%を占めており、2 番目にサンプル数の多い調査のサンプル数の倍以上になっていることが分かった。図 2 の“1 つの調査ずつ除いていく分析”は、Ritter-00 調査が平均効果量にかなりのマイナス影響を与えていたことを示しており、この調査を引込んだ結果平均効果量が 0.30 に上昇したことを示している。しかも、Ritter-00 が評価対象にしたプログラムは、学力向上への重点の置き方がそれほど強くない点で他のプログラムと異なっていた。調査の平均効果量に対する不均衡な影響力と Ritter-00 が評価対象にしているプログラムの特異性を考慮して、我々はその後の分野別・下位グループ分析からは同調査を除外することに決定した。図 3 は Ritter-00 調査を除外した、24 の調査セットの森林プロット(forest plot)と平均読解力効果量 0.30 を示している。この効果量に 95%の信頼区間(confidence interval)を適用すると、ボランティアによる個別指導の読解力成績に及ぼす平均効果は 0.18 から 0.42 の間にあることが分かる。異質性テストからは、統計的に有意とは言えない Q 値(Q = 17.29, p = .80)が出た；従って、その後の結果の報告はすべて固定効果モデルを使って行っている。

次に、我々はボランティアによる個別指導プログラムの以下の具体的な学力分野(上記方法セクション記載と同)に対する効果を調べた： 全体的読解力、文字・単語読解力、読解力、読解力オーラル・フルーエンシー、作文力、及び全体的算数力。

全体的読解力(Reading Global)

全体的読解力分野におけるアウトカム指標評価を行った調査は 13 件で、これらの調査では 195 人の個別指導を受けた生徒が分析の対象となった。全体的読解力の森林プロットが示すように(セクション 10 の図 4 参照)、8 調査がプラスの効果量を示し、6 調査がマイナスの効果量を示している。が、最も肯定的な効果量のうちの 2 つだけが統計的にゼロではない。全体としては、この成績分野の平均効果量は+0.26、即ち、統計的に有意な結果である。

文字・単語読解力(Reading Letters and Words)

文字・単語読解力分野におけるアウトカム指標評価を行った調査は 15 件で、これらの調査では 403 人の個別指導を受けた生徒が分析の対象となった。この結果の森林プロットが示すように(セクション 10 の図 5 参照)、2 調査を除く全調査がプラスの効果量を示した。これらのプラス効果量のうちの 3 つだけが自力で統計的にゼロではないが、すべての結果をプールすると、効果量は+0.41、即ち、統計的に有意な結果である。

読解力(Reading Comprehension)

読解力分野におけるアウトカム指標評価を行った調査は 8 件で、これらの調査では 293 人の個別指導を受けた生徒が分析の対象となった。読解力の森林プロットが示すように(セクシ

ン 10 の図 6 参照)、5 調査がプラスの効果量を示す一方で、3 調査のみがマイナスの効果量を示した。これらのプラス効果量のうちの 2 つだけが統計的にゼロではなく、全体的効果量は +0.18、即ち、統計的に有意とは言えない結果である。

読解力オーラル・フルーエンシー(Reading Oral Fluency)

12 件の調査が文字・単語読解力分野内でアウトカム指標評価を行った。これらの調査では 336 人の個別指導を受けた生徒が分析の対象となった。読解力の森林プロットが示すように(セクション 10 の図 7 参照)、12 調査中 10 調査がプラスの効果量を示している。これらのプラス効果量のうちの 2 つだけが自力で統計的にゼロではないが、この成績のプールされた効果量は +0.30、即ち、統計的に有意な結果である。

作文力(Writing)

レビューワが作文力分野に分類したアウトカム指標評価を行った調査は 6 件のみで、これらの調査では 111 人の個別指導を受けた生徒が分析の対象となった。作文力の森林プロットが示すように(セクション 10 の図 8 参照)、全 6 調査がプラスの効果量を示している。これらのプラス効果量のうちの 1 つだけが自力で統計的にゼロではないが、すべての結果をプールした効果量は +0.45、即ち、統計的に有意な結果である。

全体的算数力(Mathematics Global)

レビューワが全体的算数力分野に分類したアウトカム指標評価を行った調査は 5 件のみであったが、これらの調査はサンプル数の多い学位論文であった関係で、292 人の個別指導を受けた生徒が分析の対象となった。全体的算数力の森林プロットが示すように(セクション 10 の図 9 参照)、3 調査がプラスの効果量を示し、2 調査がマイナスの効果量を示した。プラス効果量のうち、統計的にゼロではないのは 2 つだけ。これら 5 調査をプールすると、この成績分野での全体的効果量は +0.27、即ち、統計的に有意とは言えない結果である。

表 2： 全調査結果の効果量の平均・標準誤差

	調査数	分析の対象となった個別指導受講生数	効果量	95% 信頼区間
Reading – Overall 読解力---総合	24	550	0.30*	(.18, .42)
Reading – Global 全体的読解力	13	195	0.26*	(.05, .48)
Reading – Letters & Words 文字・単語読解力	15	403	0.41*	(.27, .56)
Reading – Comprehension 読解力	8	293	0.18	(-.06, .42)
Reading – Oral Fluency 読解力-オーラル・フルーエンシー	12	336	0.30*	(.14, .46)
Writing 作文力	6	111	0.45*	(.19, .71)
Mathematics 算数力	5	292	0.27	(-.18, .72)

* 顕著にゼロではなく、対照群に比べ処遇グループの優位性を示している、 $p < .05$ 。

読解力アウトカムに関する調査の下位グループに対する影響分析

次に、我々は種類の異なるボランティアによる個別指導プログラム間の読解力アウトカムに対する効果差の可能性を調べた。本レビューでは各下位グループにつき少なくとも 3 件の調査が関わっている“下位グループ”効果についてのみ取り扱う。レビューの対象となった下位グループは上述の通りであるが、指導員の種類(タイプ)、生徒の学年、プログラムの構造、及び出版物の種類である。

指導員の種類による効果量の顕著な差を示すアウトカムは皆無であった。つまり、プログラムが利用している指導員が保護者、大学生、あるいは地域社会の住民の誰であれ、効果が大幅に変わるということとはなかった。

同様に、1 年生を対象にしたプログラムがそれより高学年の生徒を対象にしたプログラムより効果が顕著に異なることもなかった。我々が発見した唯一の顕著な下位グループ間差は、高構造化プログラムが低構造化プログラムに対する明らかな優位性で、効果量は高構造化プログラムが.59、低構造化プログラムが.14 となったことである。その他の読解力の成績には構造化度の高低による顕著な差はなかった。但し、全体的読解力の個別指導プログラムを対象にしていた調査で、高構造化プログラムに分類できるものを扱っていた調査は 3 件にすぎず、しかもその 3 調査とも同じ主任研究者が率いるグループが発表したものであったことに注意する必要がある(Vadasy et al., 1997a, 1997b, 2000)。

出版バイアスの可能性を評価するために、我々は“trim and fill”法を使って、大きすぎる調査を剪定し、小さな、不足気味の調査に補充する作業を行った。図 10 は、24 調査が漏斗プロットの形にきちんと対称的に適合しており、我々の結果には出版バイアスによる大した影響はないことを示唆していることを示している。さらに付け加えると、剪定と補充の作業が行われたが、過度に大きすぎると言えるほどの剪定すべき調査はなかった。従って、観察された総合効果量がバイアスのない調査に基づいている可能性大である。

出版バイアスもしくは“お蔵入りバイアス(file-drawer bias)”の可能性を調べるためのさらにもう一つのテストとして、我々は学位論文やその他出版されていない作品から選ばれた調査との区別を行った。本メタ分析に現れた全体的傾向は出版バイアスを示してはいない。つまり、本レビューで調べた各読解力分野において、雑誌に出版された調査からの効果量が学位論文や非出版調査からのものに比べて顕著に大きいということとはなかった。

表 3： 出版タイプ別効果量の平均及び標準誤差

	雑誌論文	学位論文及び非出版論文
読解力総合 (n=24)		
効果量	0.36	0.18
信頼区間	.21 から .51	-.02 から .39
調査数	14	10
全体的読解力 (n=13)		
効果量	0.43	0.13
信頼区間	.05 to .82	-.13 to .39
調査数	6	7
文字・単語読解力 (n=15)		
効果量	0.43	0.32
信頼区間	.28 to .59	-.07 to .71
調査数	13	2
読解力 (n=8)		
効果量	0.07	0.37
信頼区間	-.28 to .42	.06 to .68
調査数	5	3
読解力オーラル・フルーエンシー (n=12)		
効果量	0.31	0.12
信頼区間	.15 to .48	-.58 to .81
調査数	10	2

*顕著にゼロではなく、対照群に比べ処遇グループの優位性を示している、 $p < .05$.

5. 論考と結論

本系統的レビューの目的は、ボランティアによる個別指導プログラムの有効性に関する実験に基づく研究を集め、要約し、統合することであった。現行のエビデンスとしては、他の教育的介入に比較すると、本レビューが 20 件の無作為実地試験を掘り当てたという点で、強力である。が、この朗報には、これらの調査のほとんどが使っているサンプル数が少ないという欠点もある。本メタ分析の 28 件の調査コーホート中 19 件が対象とした個別指導実施グループの生徒数は 25 人あるいはそれ以下で；(処遇グループと対照群を合わせ)100 人を超す十分な調査サンプル数を備えていた調査コーホートは 3 件しかなかった。結局、無作為実地試験についての報告を行っているこれら 21 論文の 28 調査コーホートに含まれた全調査サンプル数は 1,676 人(873 人の個別指導の実施対象生+803 人の対照群生)となった。このように、サンプル数が少なくプログラムの効果を探知する能力に欠ける調査が数多く混じっている場合、メタ分析を行えば小型調査の結果がプールされ統計的能力が向上するという強みが発揮できる。

ボランティアによる個別指導プログラムの効果を調べるのに実験的デザインを使用した 21 論文では、多くの学力アウトカムが評価された。まず、我々はボランティアによる個別指導プログラムの全読解力アウトカムに対する全体的効果の分析を行い、0.30 標準偏差という肯定的で、統計的に有意なプラス効果を見いだした。次に、我々は個々のアウトカムを 6 つの分野にグループ分けした。うち 2 分野は、幅広く、全般的読解力と算数力の標準化された評価法を使っている、全体的読解力と全体的算数力であった。残りの 4 分野はより幅の狭い、読解力と言語能力に関連する具体的な下位能力である、文字と単語、読解力、読解力オーラル・フルーエンシー、及び作文

力であった。

中心となる目標はボランティアによる個別指導介入が低学年の学力向上に効果的な方法となり得るか否かを調べることであった。答えは、無作為実地試験の手持ちセットによると、条件付きの yes である。ボランティアによる個別指導プログラムに参加すると、総合読解力の指標が標準偏差の約 3 分の 1 上昇する。個別下位能力に関しては、ボランティアによる指導員が付き添っている生徒は、文字・単語、オーラル・フルーエンシー及び作文力に関連する評価で、指導員の付き添いのない同学年生に比べより高い得点を獲得する可能性が大である。これらのアウトカム分野に関連する効果量は、比較的一貫性を備え、0.26 から 0.45 の間にあった(上表 2 参照)。

2 番目の目標は、特定のプログラムが著しい効果を発揮するということがあるものかどうかを評価することであった。この 2 番目の目標に関しては、レビューの結果、レビューの対象プログラムはそれぞれ独特のもの(unique)で、各調査が基盤としているサンプル数は少ないということが明らかになった。さらに、これらのプログラムは反復・再生できない程小さいものでしかない。従って、“これらのプログラムのうちで最も効果的なものはどれ?”は有意義な質問とは言えない。むしろ、“効果的なプログラムの特徴とはどのようなものか?”の方がより重要な質問である。

この質問に答えるため、レビューは、1) 指導員のタイプ； 2) 生徒の年齢； 及び 3) プログラムが高構造化されているものか否か、等のプログラムの持つ各種特性に基づき微分平均効果量を算定した。これらの下位グループ分析は、総合読解力と 4 つの個別読解力分野についてのみ実施した。

が、おおむね、これらの下位グループ間に顕著な効果差は現れなかった。とはいえ、これらの分析の対象となったプログラムの特性から得られる教訓がいくつかある。本レビューの対象となった調査の大半は小学校児童(primary age students)向けの読解力中心プログラムの評価を行っていた。プログラムに肯定的効果を持たせるために高度に構造化する必要性も、指導員として特定のタイプの人物を採用する必要性もなかった。

結局のところ、本分析結果はボランティアによる個別指導を低学年児童の学力向上の方法として使う是非を決定しようとしている政策立案者や教育者が重要な証拠物件として十分利用できるものとなる筈である。

全国の教育者達が州政府の説明責任 (accountability) 制度が設けている年間達成目標を満たそうと奔走し、学力が年間達成目標に到達できない可能性のある生徒に追加サービスを無理なく提供できる方法を模索している現在、構造化の進んだ、読解力中心のボランティアによる個別指導プログラムは、読解力と言語能力の向上のための方法として考慮するに値するものであると言える。

ボランティアによる個別指導の調査は、ほとんどがサンプル数の少ない調査で占められているものの、まさに実験的デザインを利用している調査が圧倒的に多いという理由により、有用であるということも銘記に値する。これはメタ分析が持つ底力と効用を例証している。個々の調査は、単独では意味のあるプログラム効果を示していないが、総合した効果は比較的大きく、本レビューで調べた 7 つのアウトカム分野のうち 5 つで統計的に有意なものとなった。結局のところ、デ

ータを詳しく、注意深く報告している小さな無作為実地試験によって、該当分野のエビデンスの基盤が充実するのである。

本分析結果が明らかにしているもう一つのことは、この種の介入が算数の成績の向上にどの程度効果を持ちうるかに関する我々の知識はごく僅かでしかないということである。小中学生の数的思考能力の早期育成が、その後の数学の学力到達度に及ぼす重要な役割を考えると、これは嘆かわしいことである。この情報不足状況を考慮すると、早期の算数力育成に焦点を当てたボランティアによる個別指導プログラムを教育者達が開発し、実施する一方で、研究者達が協力して、これらのプログラムの有効性を評価することになれば好都合であろう。結局のところ、特定の種類の介入が早い段階から算数に悩んでいる児童に実際に役に立つか否かについて、より理解がすすめば、教育の実践者と研究者の両者共に多く資することであろう。

6. 本レビューの更新計画

本レビューの更新は2年に1度を目安にして行うことにする。できるだけ包括的な更新が行えるよう、レビューワは、より綿密に灰色文献(grey literature=政府の発行する報告書等の非営利的に出版される文献)をチェックし、ボランティアによる個別指導の分野の専門家とより広く連絡を取り、米国の国外で行われた調査も含める計画をしている。レビューワは又準実験的デザインを使用している調査を除外することにした我々の決定を見直しすることも考えている。

7. 謝辞

ボランティアによる個別指導の分野に踏み込んだ最初の調査は、1998-2000年の間ペンシルバニア大学の教員とスタッフが実施した“フィラデルフィア西部個別指導プロジェクト”の支援の下で行われた。スミス・リチャードソン基金(Smith Richardson Foundation)がボランティアによる個別指導の有効性に関する本レビューを開始するための資金を提供した。試験的レビュー用のプロトコル草案は、レベッカ・メイナード博士(Dr. Rebecca Maynard)の率いる、ペンシルバニア大学の教育学大学院の系統的レビューとメタ分析の実施に関するセミナー参加者を含む多くの同僚達の助けを得て完成された。

系統的レビューはアーカンソー大学で、キャンベル共同計画(Campbell Collaboration)の資金援助を受けて完成に至った。但し、本調査の結果は必ずしもスミス・リチャードソン基金、ペンシルバニア大学、もしくはアーカンソー大学の公式見解や政策を代表するものではない。

8. 潜在的利益相反

リッター調査 2000(Ritter 2000 study)を含めるべきか否かの決定にはゲーリー・リッター(Gary Ritter)は参加しなかった。

9. 本レビューの対象となった調査： 特性と参考事項

表 4： 本レビューの対象となった調査^{*3}の特性の要約

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
1	Allor 2004.1 (雑誌) 早期読み書き能力指導プログラム (南部の都市学校)	Gr (学年)1 N(人数) = 86 (T(処遇グループ)=61, C(対照群)=25)	教育学専攻学生 (無給) もしくは America Reads のメンバー (給付金)	週3-4回の15--20分授業を 6ヶ月	Yes	Yes	Rd-Decoding WRMT (2), TOWRE (2), DIBELS (1) Rd-Comprehension WRMT passage Rd-Authentic Oral fluency pre/post gains (前/後間の増分)	2つのコーホート年が別々に考慮された; これはコーホート1
2	Allor 2004.2 (雑誌) 早期読み書き能力指導プログラム (南部の都市学校)	Gr 1 N =157 (T=76, C=81)	教育学専攻学生 (無給) もしくは America Reads のメンバー (給付金)	週3-4回の15--20分授業を 6ヶ月	Yes	Yes	Rd-Decoding WRMT (2), TOWRE (2), DIBELS (2) Rd-Comprehension WRMT passage Rd-Authentic Oral fluency pre/post gains	2つのコーホート年が別々に考慮された; これはコーホート2
3	Baker 2000 (雑誌) SMART – Start Making a Reader Today(読書家づくり今日から始めよう)(オレゴン州)	Gr 1 then 2 (連続1-2学年) N = 84 (T=43, C=41)	地域社会(無給)	週2回の30分授業を2年間	Yes	No	Rd-Decoding WRMT word ID Rd-Comprehension WRMT word, passage (2) Rd-Authentic Oral fluency (2) Adjusted posttest (調整された事後テスト)	2年目の調査結果のみが対象

*3 これら 28 調査の全出典は本表に続く参考文献リストにあり。

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
4	Cobb 2000.1 (雑誌) 教員実習生による遊びと音韻体系認識活動 (ミッドウエスト市)	Gr 1 N = 18 (T=9, C=9)	教員実習生 (無給)	週2回の 45分授業 10週間	Yes	No	Grade 1 Rd-Global GRAT Rd-Decoding GRER (4) posttest t values (後テストt値)	3学年が異なった成績を出したので、別々に報告する；これは1学年。SDsの報告はなし、従って効果量を報告されたt値からCohen's <i>d</i> ^{*4} に変換した。
5	Cobb 2000.2 (雑誌) 教員実習生による遊びと音韻体系認識活動 (ミッドウエスト市)	Gr 2 N = 20 (T=12, C=8)	教職実習生 (無給)	週2回の 45分間授業10週間	Yes	No	Grades 2 & 3 Rd-Global GRAT Rd-Decoding GVOC Rd-Comprehension GCOMP posttest t values	3学年が異なった成績を出したので、別々に報告する；これは2学年。SDsの報告はなし、従って効果量を報告されたt値からCohen's <i>d</i> に変換した。
6	Cobb 2000.3 (雑誌) 教員実習生による遊びと音韻体系認識活動 (ミッドウエスト市)	Gr 3 N = 18 (T=9, C=9)	教職実習生 (無給)	週2回の 45分間授業10週間	Yes	No	Grades 2 & 3 Rd-Global GRAT Rd-Decoding GVOC Rd-Comprehension GCOMP posttest t values	3学年が異なった成績を出したので、別々に報告する；これは3学年。SDsの報告はなし、従って効果量を報告されたt値からCohen's <i>d</i> に変換した。

^{*4} $d = t \sqrt{1/nE + 1/nC}$ (Glass, McGaw, & Smith, 1981)

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
7	Cook 2001.1 (学位論文) 最小限の研修を受けた指導員が America Readsを教材として使用(アリゾナ州フェニックス近郊)	Gr 1 N = 26 (T=12, C=14)	大学生 (一部無給、一部労働学習)	週2回の 45分 間授業7ヶ月 間	Yes	No	Rd-Global WRAT3 posttest	指導員の自然減少と頼りにならない指導員の問題が発生。 学年別に報告されていたので、データは3つの個別コーホートとして分析した。これは1学年。
8	Cook 2001.2 (学位論文) 最小限の研修を受けた指導員が America Readsの教材を使用(アリゾナ州フェニックス近郊)	Gr 2 N = 17 (T=7, C=10)	大学生 (一部無給、一部労働学習)	週2回の 45分 間授業7ヶ月 間	Yes	No	Rd-Global WRAT3 posttest	指導員の自然減少と頼りにならない指導員の問題が発生。 学年別に報告されていたので、データは3つの個別コーホートとして分析した。これは2学年。
9	Cook 2001.3 (学位論文) 最小限の研修を受けた指導員が America Readsを教材として使用(アリゾナ州フェニックス近郊)	Gr 3 N = 17 (T=11, C=6)	大学生 (一部無給、一部労働学習)	週2回の 45分 間授業7ヶ月 間	Yes	No	Rd-Global WRAT3 posttest	指導員の自然減少と頼りにならない指導員の問題が発生。 学年別に報告されていたので、データは3つの個別コーホートとして分析した。これは3学年。

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
10	Erion 1994 (学位論文) 保護者がフラッシュカードや読本を使って指導 (ペンシルバニア州北西部農村地)	Gr 2 N = 24 (T=12, C=12)	保護者(無給)	週5回の 15分 授業 6週間	Yes	No	Rd-Authentic oral reading fluency pre/post gains	
11	Mahoney 1986 (学位論文) 保護者が算数を指導 (オハイオ州レイクウッド)	Gr 3 N = 150 (T=75, C=75)	保護者(無給)	週5回の 30分 授業4週間	No	Yes	算数 50項目の掛け算テスト; 後テスト	生徒の無作為振り分けによる教室構成; テストは標準化されたものではなかった。
12	Mayfield 2000 (学位論文) Edmark Reading Program(エドマーク読解プログラム) – 構造化された個別指導 (ルイジアナ州北部農村地)	Gr 1 N = 62 (T=31, C=31)	America Reads の メンバー	週5回の 15分 授業(1学年2 学期制の)1学 期間	Yes	Yes	Rd-Decoding WRMT (3) Rd-Comprehension WRMT passage pre/post gains (前/後増分) [後分についても調整済み]	一つの読解指標 (word attack) がSD-Tの間違った値を示していたので、効果量の計算はブールされたSDではなくSD-Cを使って行った。正式な指標はカリキュラム(教科課程)とのつながりが強すぎたので除外した。
13	McKinney 1995 (学位論文) Leap Frog Program(カエル跳びプログラム): 教会における放課後個別指導プログラム (ミシシッピー州北東部農村地)	Gr 1,2 N = 44 (T=20, C=24)	大学生(無給)	週4回の1時間 授業 22週間	No	No	Rd-Global Stanford-Reading Mathematics Stanford-Mathematics pre/post gains	成績は百分順位として報告された。

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
14	Mehran 1988 (雑誌) Reading Made Easy 保護者による個別指導(西部の小都市)	Gr 1 N = 76 (T=38, C=38)	保護者 (無給)	週3回の 15分授業 1学年間	Yes	Yes	Rd-Global WJPEB, CTBS Rd-Decoding WJPEB (2), CTBS (2), Harrison (4) Rd-Comprehension WJPEB, CTBS posttest	処遇の忠実性- 保護者の大部分の授業出勤数は月4回。 前テストを実施したのはCTBSのみ、但し後テストと同じ下位得点(subscores)を使っていなかったため、増分(gains)は計算されなかった。
15	Miller 1994 (学位論文) Paired Reading (ペア読書)保護者が指導(中西部地区)	Gr 2-4 N = 52 (T=26, C=26)	保護者 (無給)	週4回の 10-15分授業10週間	Yes	No	Rd-Global GORT-D, Pre/post gain scores(前/後増分点数)	処遇の忠実性-保護者は授業をテープに録画することになっていたが、録画した人はごく少数。
16	Morris 1990.1 (雑誌) Howard Street Tutoring Model(ハワード通り個別指導モデル); 放課後プログラム (イリノイ州)	Gr 2-3 N = 34 (T=17, C=17)	地域社会ボランティア、年齢にばらつきあり (無給)	週4回の30分授業 1学年全体	Yes	Yes	Rd-Decoding Word recognition measures (3) Rd-Authentic. Basal passages Writing Spelling correct, qualitative Gain scores only* ⁵ (増分点数のみ5)	1986-87年と1987-88年コーホートは本メタ分析においては別々の調査。それぞれのコーホートには、幾人か同一の生徒が混じっている。 指標は規格化されていない可能性あり。

*⁵ この調査は増分得点とその標準偏差についてのみ報告しており、前/後の相関関係には触れていない。従って、後テストのプールされた標準偏差を引き出すことはできなかった。我々の効果量計算用に後テストのプール標準偏差を見積もるために、我々は、0.5 の前/後の相関関係とそれに等しい標準偏差を想定する増分得点の標準偏差を使った。

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
17	Morris 1990.2 (雑誌) Howard Street Tutoring Model(ハワード通り個別指導モデル); 放課後プログラム (イリノイ州)	Gr 2-3 N = 26 (T=13, C=13)	地域社会ボランティア、年齢にばらつきあり (無給)	週4回の30分授業 1学年全体	Yes	Yes	Rd-Decoding Word recognition measures (3) Rd-Authentic. Basal passages Writing Spelling correct, qualitative Gain scores only	1986-87年と1987-88年コーホートは本メタ分析においては別々の調査。それぞれのコーホートには、幾人か同一の生徒が混じっている。 指標は規格化されていない可能性あり。
18	Nielson 1991 (学位論文) 保護者及び成人の読解力におけるボランティアによる個別指導 (ユタ州デルタの農村小学校)	Gr 3 N=43 (T=29, C=14)	保護者、成人ボランティアによる (無給)	週あたりの授業回数と時間の記述なし; プログラムの実施期間は 9ヶ月	Yes	Yes	Rd-Comprehension Stanford Achievement Reading Comprehension posttest	フォローアップ調査結果は含まれていなかった。保護者と成人ボランティアによるは分析のために共に一つの処遇グループに入れられた。
19	Parham 1994.1 (学位論文) 訓練を受けた指導員による代数学入門コンセプトについての登校前個別指導(郊外地校)	Gr 7 N = 32 (T=16, C=16)	地域社会ボランティア(無給)	週1回の 60分授業5週間	No	Yes	Mathematics OHAPT posttest	我々は成人指導員に関する調査結果だけに絞る、同年輩指導員は除外した。午前7時という開始時間に拘わらず指導員の自然減に関する記述はなし。単独の対照群が、2つのパーハム (Parham) 処遇グループの間で半分ずつに分けられた。

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
20	Parham 1994.2 (学位論文) 訓練を受けていない指導員による代数学入門コンセプトについての登校前個別指導(郊外地校)	Gr 7 N = 32 (T=16, C=16)	地域社会ボランティア(無給)	週1回の 60分 授業5週間	No	No	Mathematics OHAPT posttest	我々は成人指導員に関する調査結果だけに絞り、同年輩指導員は除外した。午前7時という開始時間に拘わらず指導員の自然減に関する記述はなし。単独の対照群が、2つのパーハム (Parham) 処遇グループの間で半分ずつに分けられた。
21	Powell-Smith 2000 (雑誌) 保護者の児童文学もしくは基礎読本を使っの指導：実習訓練、フィードバック(批評)及び処遇の忠実性の監視を行う案内人付き	Gr 2 N = 36 (T=24, C=12)	保護者(無給)	週4回の 20分 授業 5週間	Yes	Yes	Rd-Authentic Curriculum Based Measure (CBM), Test of Reading Fluency (TORF) Pre/post gain scores	個別指導の2つのバージョン、即ち基礎読本を利用するバージョンと児童文学を利用するバージョンが合体された。若干の処遇の忠実性に関する危惧有り。
22	Pullen 2004 (雑誌) 読解及び新しい本の読み方指導を行いながらの反復読書(フロリダ州北方中央部)	Gr 1 N = 49 (T=25, C=24)	大学生(無給)	12週間中合計 40回の15分授業	Yes	Yes	Rd-Decoding Jump Start (3) pre/post gains, WDRB (2) posttest	

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
23	Rimm-Kaufman 1999 (雑誌) 意味を理解しながら読むことの重要性を強調する包括的読み方モデル (マサチューセッツ州ケンブリッジ)	Gr 1 N = 42 (T=21, C=21)	地域社会ボランティア(無給)	週3回 45分授業 8ヶ月間	Yes	Yes	Rd-Decoding Observational Survey (観察調査)(3) Rd-Writing Observational Survey (2) Rd-Authentic Observational Survey pre/post gains	
24	Ritter 2000 (学位論文) 西フィラデルフィア個別指導プロジェクト; 大学を中心とするパートナーシップ (ペンシルバニア州フィラデルフィア)	Gr 2-5 N = 385 (T=196, C=189)	大学生 (無給)	週1時間 1学年全期間	No	No	Rd-Global (SAT-9) Mathematics (SAT-9) posttest	(指導員の自然減後の) 成績の下位グループ別サンプル数は報告に無かったが、T/C(処遇グループ対対照群)の総サンプル数比から見積もられた。
25	Vadasy 1997a (雑誌) Sound Partners(健全なパートナー); 音韻体系認識、単語ID、文章の読み方、作文等に関する、筋書き原稿付き100レッスン(大都市部)	Gr 1 N = 40 (T=20, C=20)	地域社会ボランティア (給付金)	週4回の 30分授業27週間	Yes	Yes	Rd-Global WRAT-R Rd-Decoding WJPEB, Dolch, Yopp-Singer Rd-Writing Writing measure, Spelling (2) Rd-Authentic Analytic Reading Inventory posttest	指導員の一貫性に若干の問題有り。

	調査 / 種類 / プログラムの説明	生徒	指導員	時間 / 持続期間	読解力中心?	高構造化されている?	アウトカム指標	危惧/コメント
26	Vadasy 1997b (雑誌) Vadasy 1997aの第2年度 (大都市部)	Gr 1 N = 40 (T=20, C=20)	地域社会ボランティア(給付金)	週4回の 30分授業27週間	Yes	Yes	Rd-Global WRAT-R Rd-Decoding WJPEB, Dolch, Bryant, Yopp-Singer Rd-Writing Writing measure, Spelling (2) Rd-Authentic Analytic Reading Inventory 調整された後テスト平均値	指導員の一貫性に若干の問題有り。 Lesson Word Listの成績は、当処遇特有のものでありすぎたので、除外した。
27	Vadasy 2000 (雑誌) Sound Partners ; 音韻体系認識、単語ID、文章の読み方、作文等に関する、筋書き原稿付き100レッスン(大都市部)	Gr 1 N=46 (T=23, C=23)	地域社会ボランティア(給付金)	週4回の 30分授業1学年期間	Yes	Yes	Rd-Global WRAT-R Rd-Decoding WJPEB, Dolch, Bryant, Yopp-Singer Rd-Writing Writing measure, Spelling (2) Rd-Authentic Analytic Reading Inventory (2) adjusted posttest means	当座の後テスト結果のみ使用。2学年になってからの追跡結果は含まず。
28	Weiss 1988 (非出版レポート) ペア読書 (郊外地区)	Gr 3-6 N = 20 (T=11, C=9)	地域社会 (無給)	週4回の 20-30分授業 11週間	Yes	No	Rd-Global BASIS Rd-Authentic CBM pre/post gains	

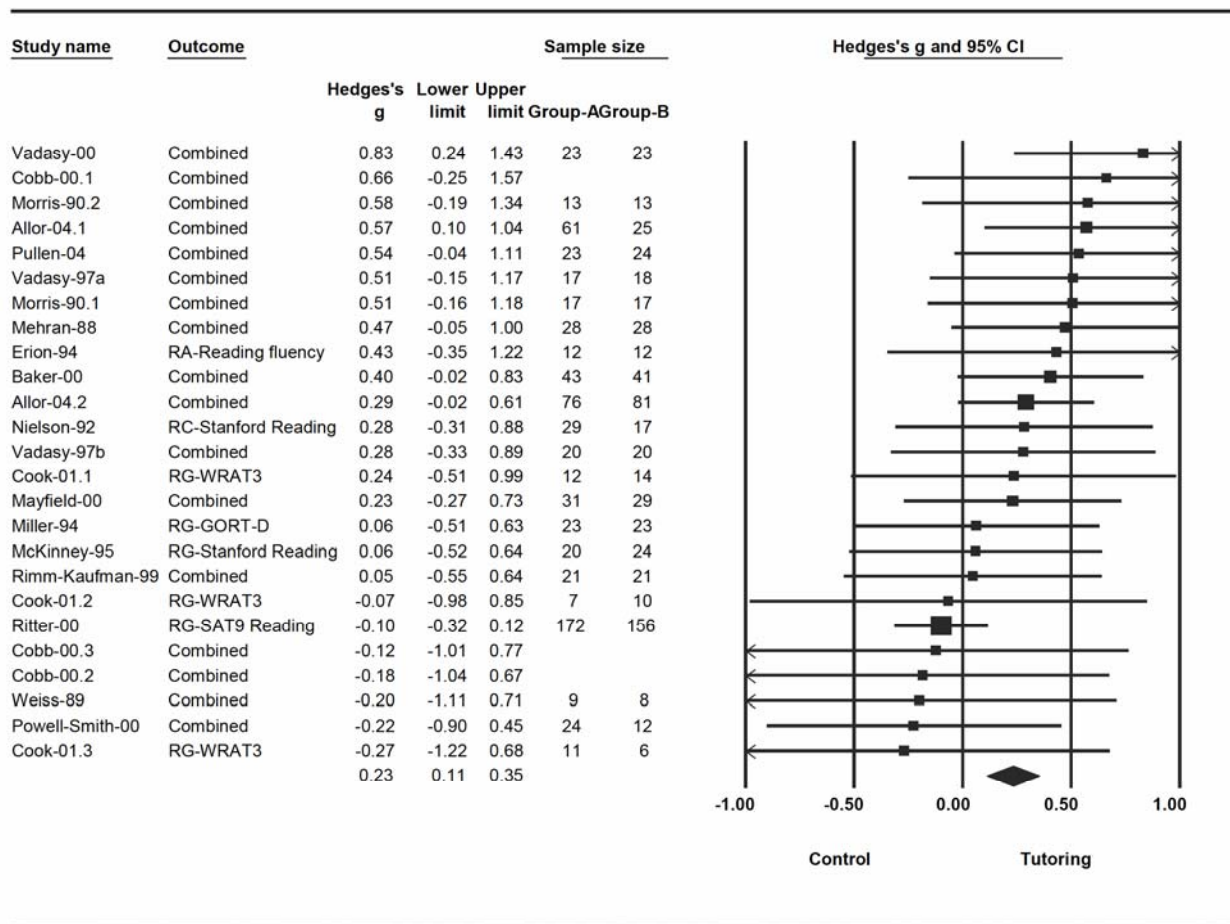
参考文献リスト — 対象となった調査 (21論文)

- Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research & Practice, 19*(2), 116-129. (studies 1 and 2)
- Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly 35*(4), 494-519. (study 3)
- Cobb, J. B. (2001). The effects of an early intervention program with preservice teachers as tutors on the reading achievement of primary grade at risk children. *Reading Horizons, 41*(3), 155-173. (studies 4, 5, and 6)
- Cook, J. A. (2002). Every moment counts: Pairing struggling young readers with minimally trained tutors (Doctoral dissertation, Arizona State University, 2001). *Dissertation Abstracts International, 62*, 8A. (studies 7, 8, and 9)
- Erion, R. J. (1994). Parent tutoring, reading instruction and curricular assessment (Doctoral dissertation, Indiana University of Pennsylvania, 1994). *Dissertation Abstracts International, 54*, 11A. (study 10)
- Mahoney, E.J. (1986). The effects of parental involvement on achievement of third-grade students (Doctoral dissertation, University of Akron, 1985). *Dissertation Abstracts International, 47*, 2A. (study 11)
- Mayfield, L. G. (2000). The effects of structured one-on-one tutoring in sight word recognition of first-grade students at-risk for reading failure (Doctoral dissertation, Louisiana Tech University, 2000). *Dissertation Abstracts International, 61*, 2A. (study 12)
- McKinney, A. D. (1995). The effects of an after school tutorial and enrichment program on the academic achievement and self-concept of below grade level first and second-grade students (Doctoral dissertation, University of Mississippi, 1995). *Dissertation Abstracts International, 56*, 6A. (study 13)
- Mehran, M., & White, K. R. (1988). Parent tutoring as a supplement to compensatory education for first-grade children. *RASE: Remedial & Special Education, 9*(3), 35-41. (study 14)
- Miller, B.V. (1995). An evaluation of the paired reading program using competency-based training. (Doctoral dissertation, University of Wisconsin - Madison, 1994) *Dissertation Abstracts International, 55*, 7A. (study 15)
- Morris, D., Shaw, B., & Perney, J. (1990). Helping low readers in grades 2 and 3: An after-school volunteer tutoring program. *Elementary School Journal, 91*(2), 133-150. (studies 16 and 17)
- Nielson, B. B. (1992). Effects of parent and volunteer tutoring on reading achievement of third grade at-risk students (Doctoral Dissertation, Brigham Young University, 1991). *Dissertation Abstracts International, 52*, 10A. (study 18)
- Parham, J. W. (1994). An analysis of the effects of tutoring on seventh grade students engaged in the mastery of pre-algebra concepts (Doctoral dissertation, University of Maryland College Park, 1993). *Dissertation Abstracts International, 54*, 11A. (study 19)
- Powell-Smith, K. A., Shinn, M. R., Stoner, G., & Good, R. H. (2000). Parent tutoring in reading using

- literature and curriculum materials: Impact on student reading achievement. *School Psychology Review*, 29(1), 5-27. (study 20)
- Pullen, P.C., Lane, H.B., & Monaghan, M.C. (2004). Effects of a volunteer tutoring model on the early literacy development of struggling first grade students. *Reading Research & Instruction*, 43(4), 21-40. (study 21)
- Rimm-Kaufman, S. E., Kagan, J., & Byers, H. (1999). The effectiveness of adult volunteer tutoring on reading among "at risk" first grade children. *Reading Research & Instruction*, 38(2), 143-152. (study 22)
- Ritter, G. W. (2000) The academic impact of volunteer tutoring in urban public elementary schools: Results of an experimental design evaluation (Doctoral dissertation, University of Pennsylvania, 2000). *Dissertation Abstracts International*, 61, 3A. (study 23)
- Vadasy P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997a). Community-based early reading intervention for at-risk first graders. *Learning Disabilities Research and Practice*, 12(1), 29-39. (study 24)
- Vadasy P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997b). The Effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly*, 20(2), 126-139. (study 25)
- Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities*, 33(6), 579-590. (study 26)
- Weiss, J.A. (1989). *Paired reading with adult volunteer tutors as a reading intervention for students with reading disabilities*. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989). (ERIC Document Reproduction Services No. ED 305606). (study 27)

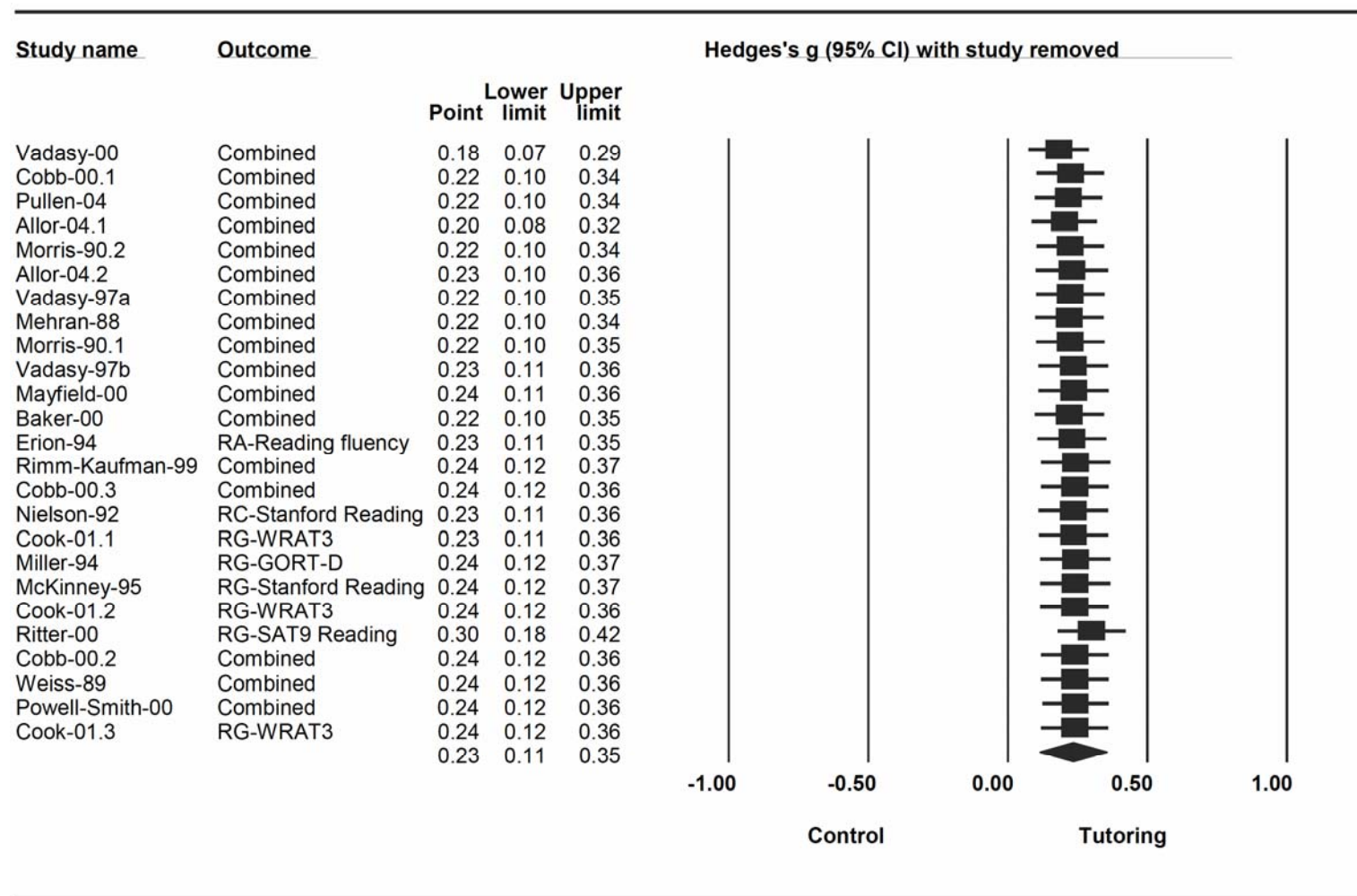
10. 図：すべての主要アウトカムの森林プロット

図 1：ボランティアによる個別指導の子供の読解力アウトカムに対する効果の森林プロット



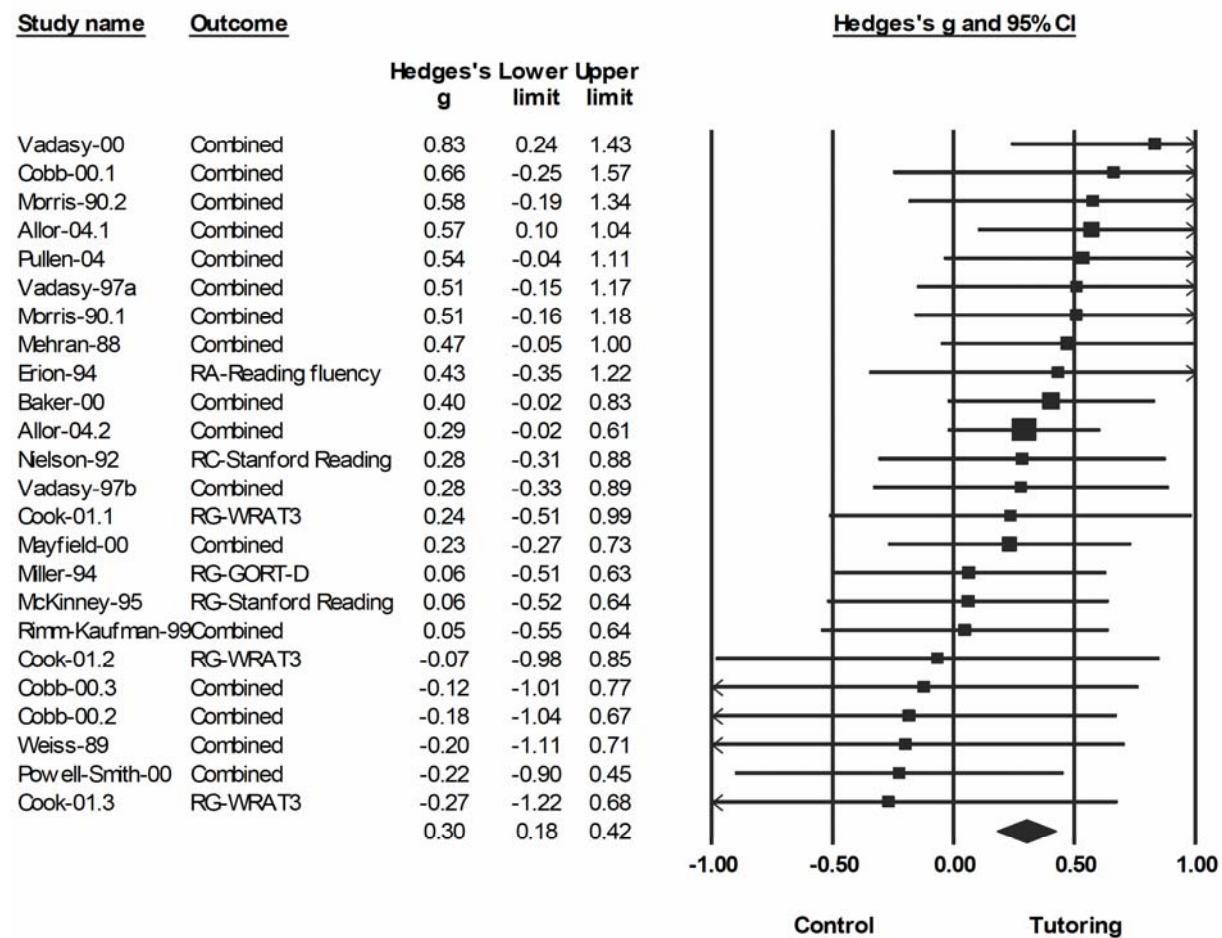
固定効果モデルの異質性統計：Q = 27.20, df = 24, p = .30, and I squared = 11.75

図 2 : 1 調査除外後のボランティアによる個別指導の子供の読解力アウトカムに対する効果の森林プロット



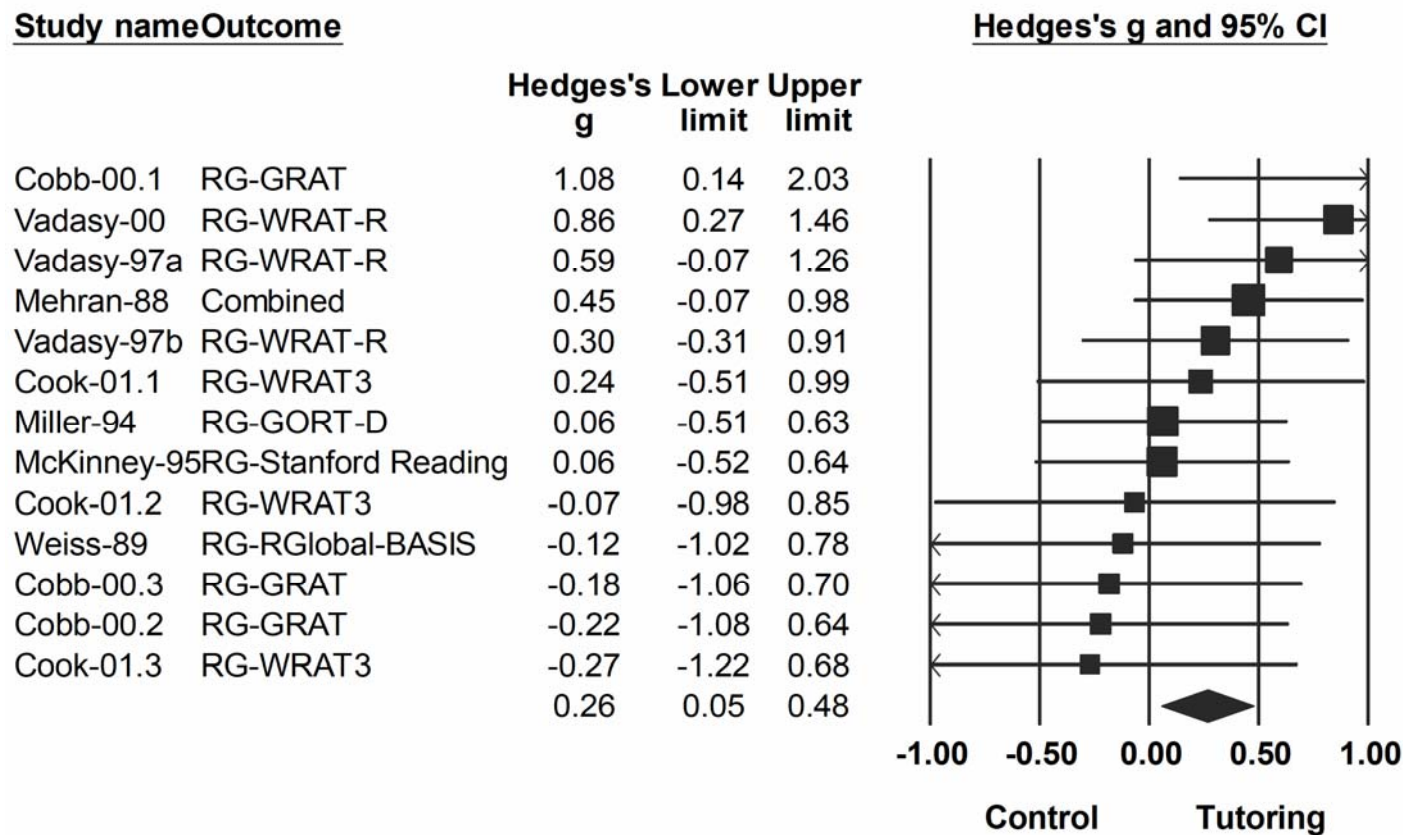
固定効果モデルの異質性統計 : $Q = 27.20$, $df = 24$, $p = .30$, and $I^2 = 11.75$

図 3 : (Ritter-00 を除外しての)ボランティアによる個別指導の子供の読解力アウトカムに対する効果の森林プロット



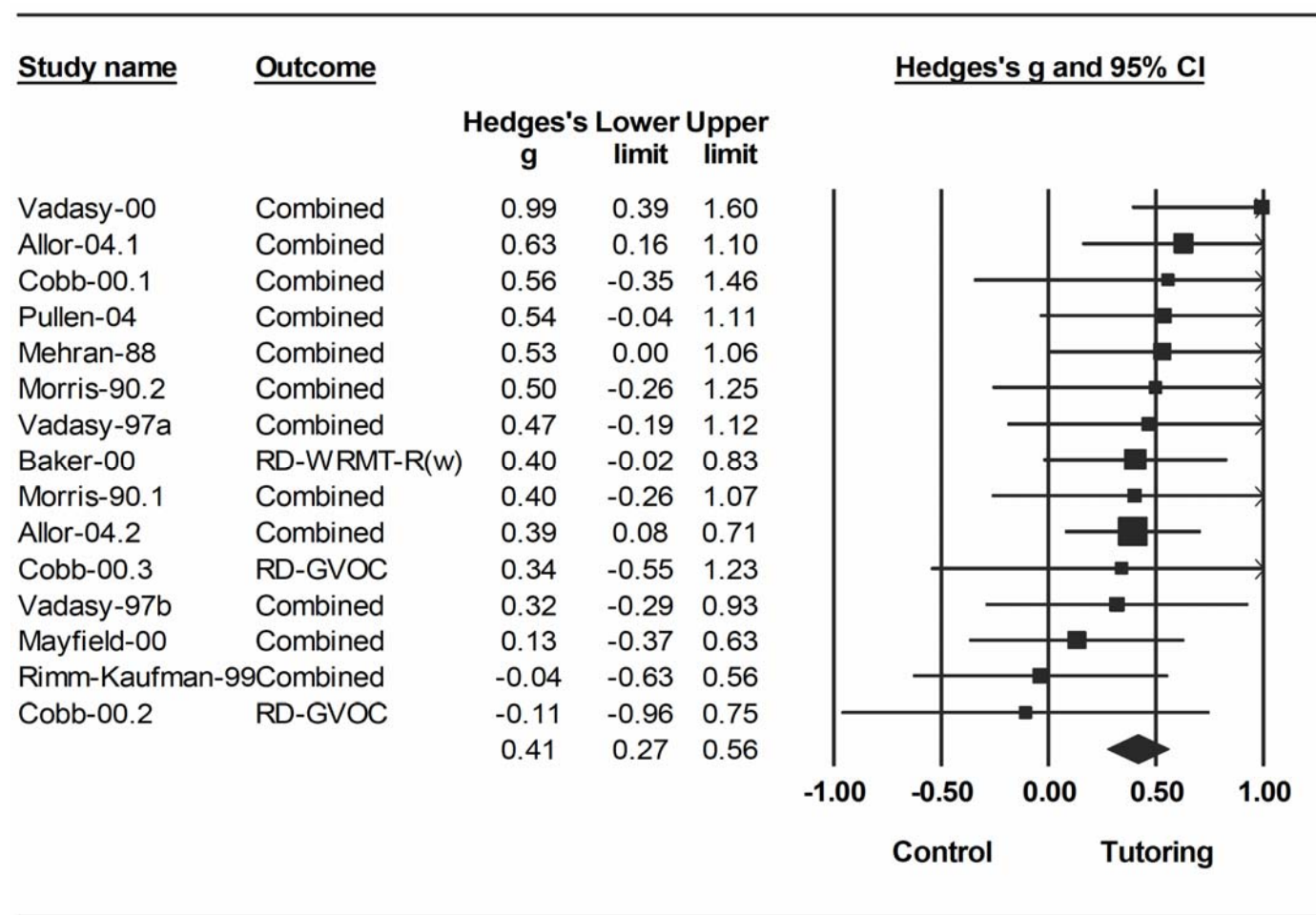
固定効果モデルの異質性統計 : $Q = 17.29$, $df = 23$, $p = .80$, and $I^2 = 0.00$

図 4 : (Ritter-00 を除外しての)全体的読解力のアウトカム指標の森林プロット



固定効果モデルの異質性統計 : $Q = 13.84$, $df = 12$, $p = .31$, and $I^2 = 13.27$

図 5 : 文字・単語読解力のアウトカム指標の森林プロット



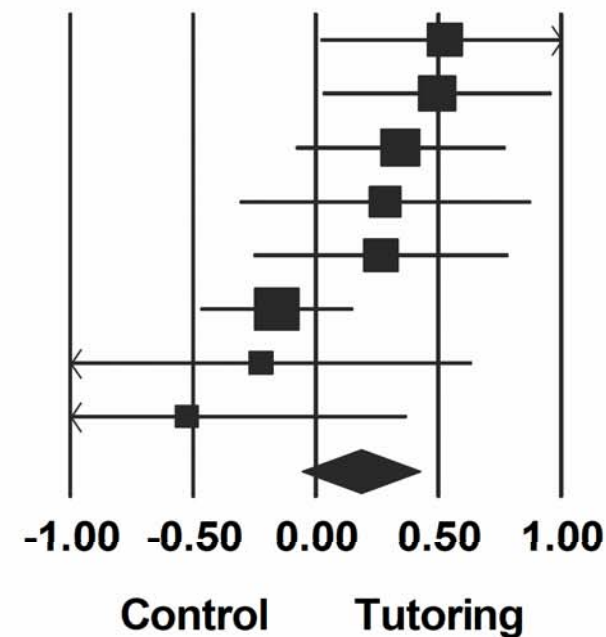
固定効果モデルの異質性統計 : $Q = 9.83$, $df = 14$, $p = .78$, and $I^2 = 0.00$

図 6 : 読解力のアウトカム指標の森林プロット

Study name Outcome

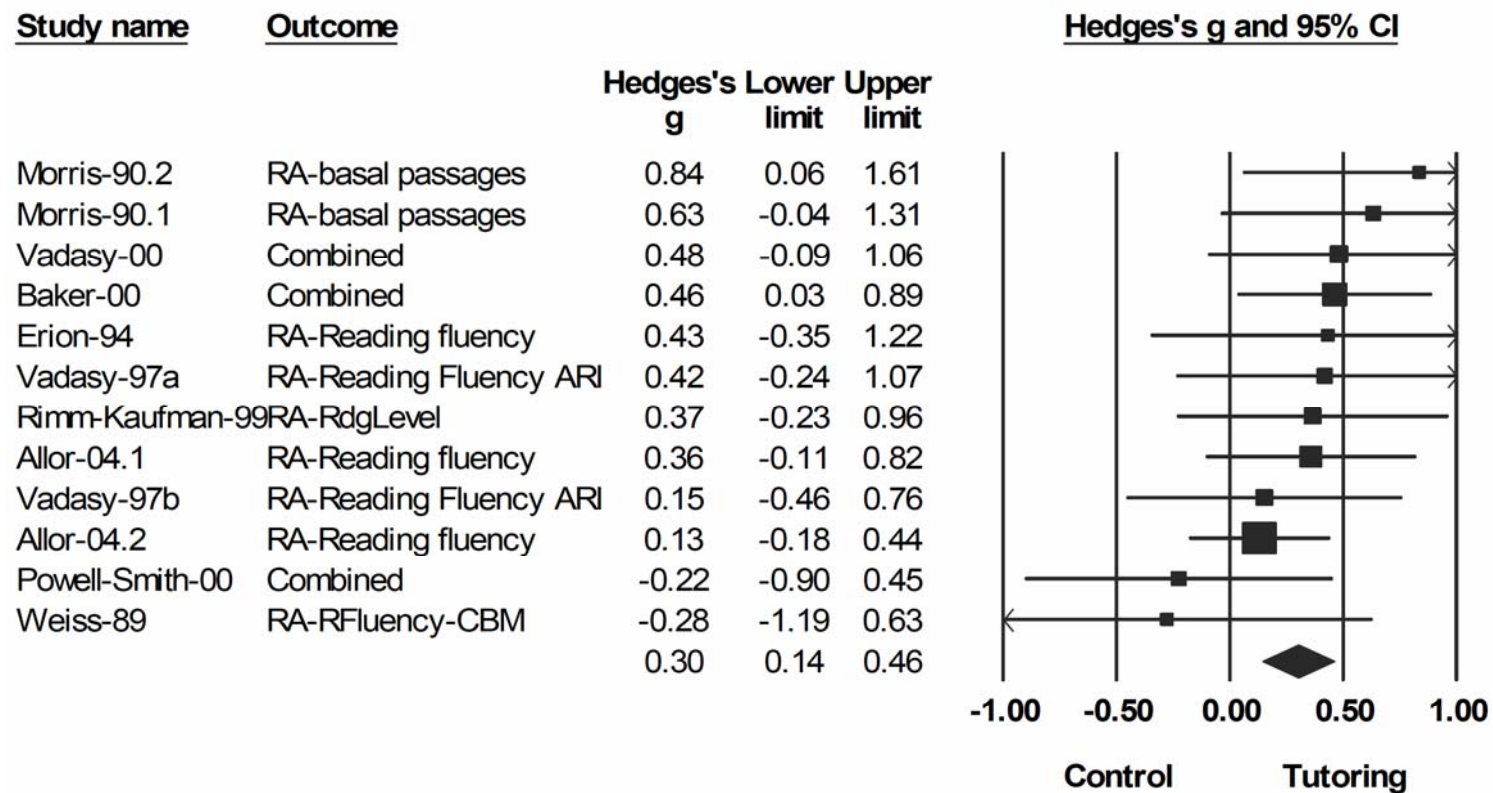
		Hedges's g	Lower limit	Upper limit
Mayfield-00	RC-WRMT-R pass comp	0.53	0.02	1.04
Allor-04.1	RC-WJR pass comp	0.50	0.03	0.96
Baker-00	Combined	0.35	-0.08	0.77
Nielson-92	RC-Stanford Reading	0.28	-0.31	0.88
Mehran-88	Combined	0.27	-0.25	0.79
Allor-04.2	RC-WJR pass comp	-0.16	-0.47	0.15
Cobb-00.2	RC-GCOMP	-0.22	-1.08	0.64
Cobb-00.3	RC-GCOMP	-0.52	-1.42	0.37
		0.18	-0.06	0.42

Hedges's g and 95% CI



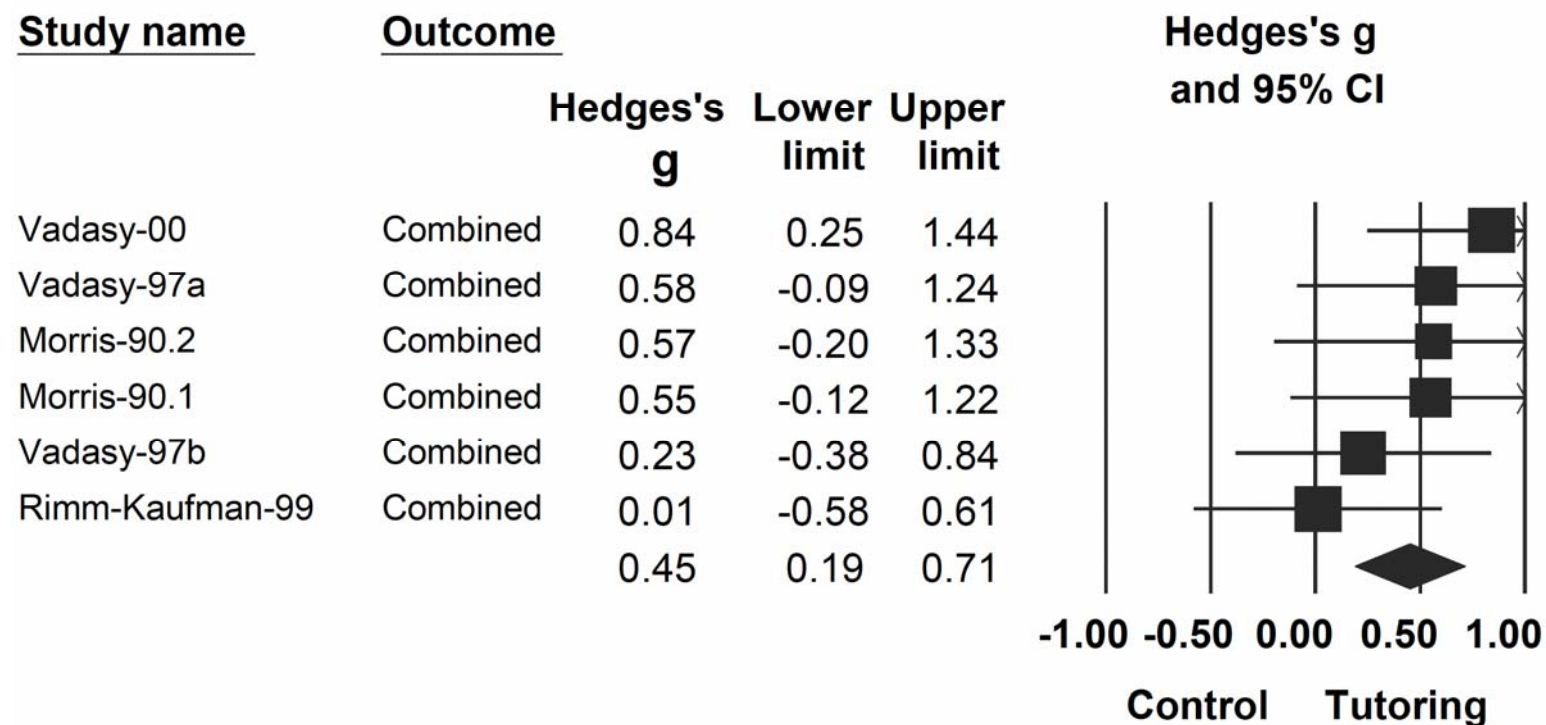
固定効果モデルの異質性統計 : $Q = 12.03$, $df = 7$, $p = .10$, and $I^2 = 41.82$

図 7 : 読解力オーラル・フルーエンシーのアウトカム指標の森林プロット



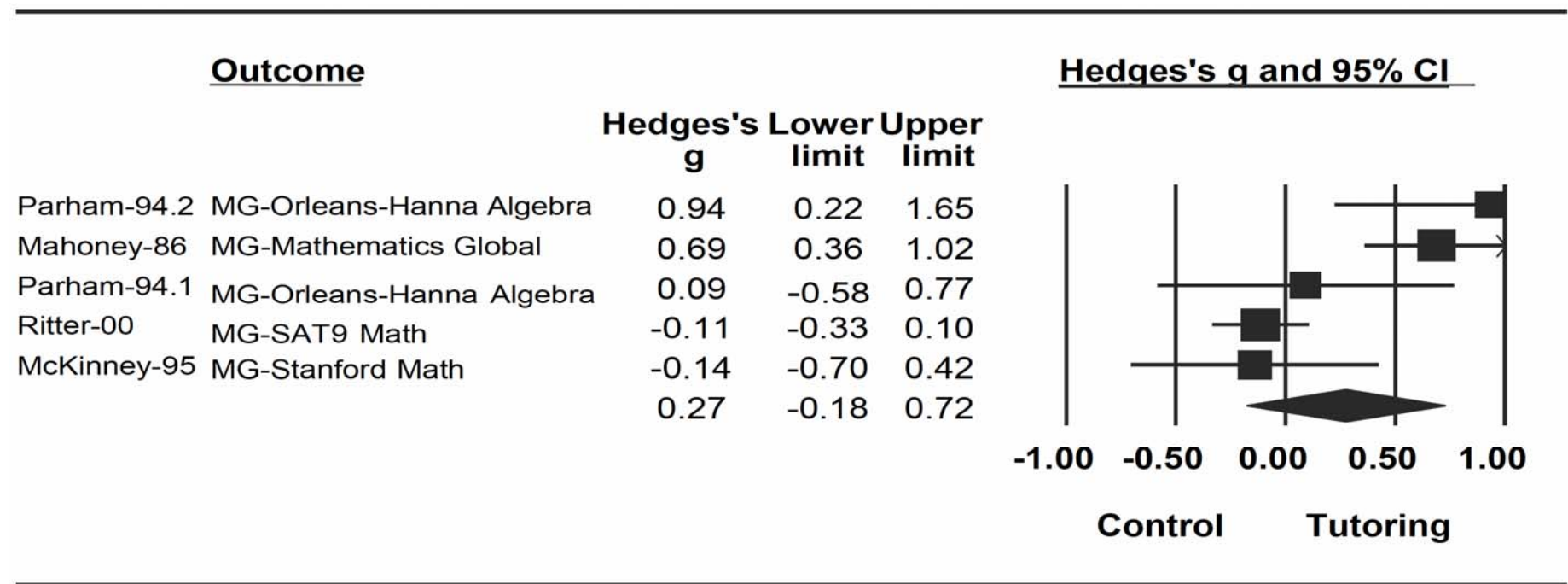
固定効果モデルの異質性統計 : $Q = 9.24$, $df = 11$, $p = .60$, and $I^2 = 0.00$

図 8 : 作文力のアウトカム指標の森林プロット



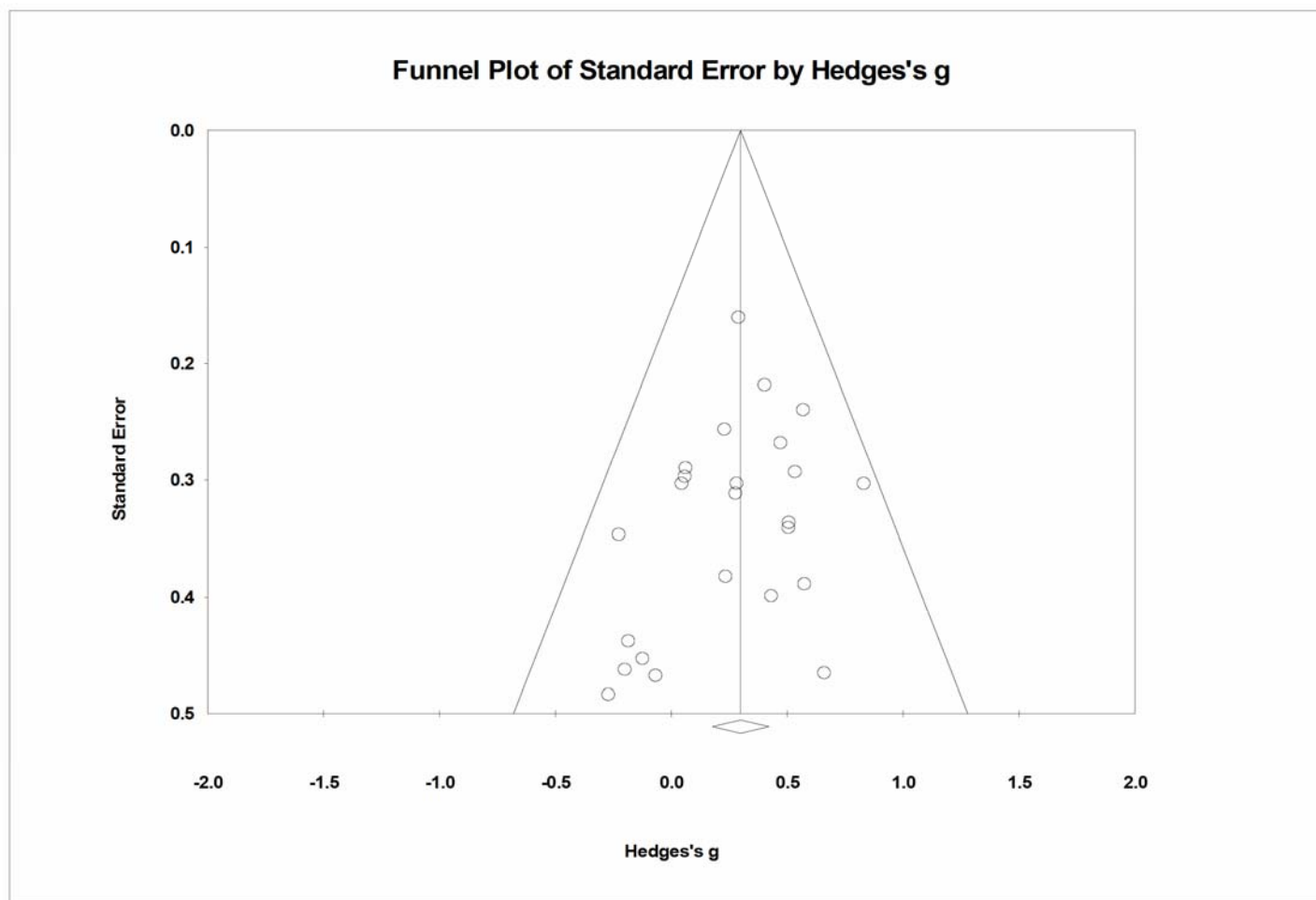
固定効果モデルの異質性統計 : $Q = 4.57$, $df = 5$, $p = .47$, and $I^2 = 0.00$

図 9 : 全体的算数力のアウトカム指標の森林プロット



固定効果モデルの異質性統計 : $Q = 21.68$, $df = 4$, $p < .01$, and $I^2 = 81.55$

図 10 : 総合読解力のヘッジズ g による標準誤差の漏斗プロット(24 調査、Ritter-00 除外)



11. 追加表

表5：各種プログラムの構造化レベル

調査名	コード	構造化度の高/低の根拠
Allor-04.1,2	高	授業にゲーム、単語学習活動、及び読書活動が含まれていた。
Baker-00	低	指導員には具体的なテクニック(技法)よりはむしろ大まかな枠組が指示されていた。
Cobb-01.1,2,3	低	小さな玩具、ゲーム及び子供向けの本が利用できるようになっていたが、各種活動は指導員が創造した。
Cook-01.1,2,3	低	指導員は America Readsの案内書を与えられていたが、そこに記載されている戦略のどれについても利用を義務づけられていなかった。
Erion-94	低	保護者は反復朗読と連続して基礎読本に目を通す作業を繰り返した。
Mahoney-86	高	保護者には自分達の子供と共に勉強する掛け算のレッスンと練習問題が与えられた。
Mayfield-00	高	15分のエドマーク読解力プログラム(Edmark Reading Program)、順序の定まった、反復練習式の、視覚語(sight-word)習得方式
McKinney-95	低	担任の先生が生徒を宿題と共に“馬跳び(Leap Frog)”放課後個別指導プログラムに送り込む。
Mehran-88	高	保護者が“易しい読み方(Reading Made Easy)”レッスンを施す。
Miller-88	低	ペア読書：保護者は子供に、単独あるいは一緒に、子供が選んだ本を読ませる。
Morris-90.1,2	高	60分の授業は綿密に計画され、読み方、単語学習及び書き方の勉強でぎっしり詰まっている。
Nielson-92	高	指導員にはストップウォッチ、フラッシュカード、物語集、及び記録簿が支給され、それらを授業でどのようにして使うかの訓練を受ける。
Parham-94.1	高	指導員は訓練と指導用教材の支給を受けた。
Parham-94.2	低	個別指導の授業が正規の授業の補習授業となった。
Powell-Smith-00	高	20分の授業は、予習、子供の音読、精選活動の3つの構成部分より成っていた。
Pullen-04	高	15分の授業はGaining Fluency(流暢さの習得)、Measuring Progress(進歩状況の測定)、及びReading a New Book(新たな本を読む)で構成されていた。
Rimm-Kaufman-99	高	各授業は一定のスケジュールに従って進められた：最初になじみのある教材、次により難しい教材、締めくくりはなじみのある作業。
Ritter-00	低	授業で扱われる主題は各学校の現場、及び個々の指導員と生徒のペア(組合せ)毎にまちまちであった。
Vadasy-97a, 97b, 00	高	健全パートナー・プログラム(Sound Partners program)：100 の放課後レッスン、各レッスンの長さ30分
Weiss-89	低	ペア読書：生徒が指導員と共に読む材料を選んだ。

表 6：調査別アウトカム分野別効果量結果の構成要素

読解力： 全体的読解力

調査	指標	Unbiased <i>d</i> (不偏 <i>d</i>)
Cobb-01.1	Gates-MacGinitie Reading Test (GRAT)	1.08
Cobb-01.2	GRAT	-0.22
Cobb-01.3	GRAT	-0.18
Cook-01.1	Wide-Range Achievement Test (WRAT)	0.24
Cook-01.2	WRAT	-0.06
Cook-01.3	WRAT	-0.27
McKinney-95	Stanford Achievement Tests (第8版)	0.06
Mehran-88	Woodcock-Johnson Psycho-Educational Battery(WJPEB) Comprehensive Test of Basic Skills (CTBS) <i>Combined mean (合算平均)</i>	0.66 0.25 0.45
Miller-88	Gray Oral Reading Test-Diagnostic (GORT-D)	0.06
Ritter-00	Stanford Achievement Test (9th ed.)	-0.10
Vadasy-97a	WRAT – Reading	0.59
Vadasy-97b	WRAT – Reading	0.30
Vadasy-00	WRAT – Reading	0.86
Weiss-89	Basic Achievement Skills Individual Screener(BASIS) – Raw score	-0.18

読解力： 文字・単語

調査	指標	Unbiased <i>d</i>
Allor-04.2	Woodcock Johnson-Revised (WJ-R) Word ID	0.57
	WJ-R Word Attack	0.85
	Test of Word Reading Efficiency (TOWRE) Real Word	0.30
	TOWRE Nonword	0.80
	Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Phoneme-segmentation fluency <i>Combined Mean</i>	0.62 0.63
Allor-04.2	WJ-R Word ID	0.20
	WJ-R Word Attack	0.66
	TOWRE Real Word	0.13
	TOWRE Nonword	0.55
	DIBELS Phoneme-segmentation fluency	0.44
	DIBELS Nonsense-word fluency <i>Combined Mean</i>	0.40 0.39
Baker-00	Woodcock Reading Mastery Test-Revised (WRMT-R) Word ID	0.40
Cobb-01.1	GRAT-Initial consonants & clusters (GRER1)	0.32
	GRAT-Final consonants & clusters (GRER2)	0.71
	GRAT-Vowels (GRER3)	0.95
	GRAT-Use of sentence context (GRER4) <i>Combined Mean</i>	0.25 0.56
Cobb-01.2	GRAT-Vocabulary (GVOC)	-0.11
Cobb-01.3	GRAT-Vocabulary (GVOC)	0.33
Mayfield-00	WRMT-R Word ID	-0.01
	WRMT-R Letter ID	0.26
	WRMT-R Word Attack <i>Combined Mean</i>	0.13 0.13

Mehran-88	WJPEB Letter/Word ID	0.56	<i>Combined Mean</i> 0.53
	WJPEB Word Attack	0.71	
	CTBS Word Analysis	0.15	
	CTBS Vocabulary	0.33	
	Harrison Criterion Referenced Test (HCRT) – Producing Sounds	0.75	
	HCRT Consonant Sound	0.62	
	HCRT Short Vowels	0.33	
	HCRT Digraphs and Combinations	0.77	
Morris-90.1	Word recognition (untimed)	0.24	<i>Combined Mean</i> 0.46
	Basal word recognition	0.68	
Morris-90.2	Word recognition (timed/flash)	0.61	<i>Combined Mean</i> 0.60
	Word recognition (untimed)	0.43	
	Basal word recognition	0.75	
Pullen-04	Jump Start – Phonological Awareness	0.39	<i>Combined Mean</i> 0.63
	Jump Start – Sight Words	0.24	
	Jump Start – Nonword Decoding	0.82	
	WDRB – Letter Word ID	0.38	
	WDRB – Word Attack	0.86	
Rimm-Kaufman-99	Observational Survey – Letters	0.09	<i>Combined Mean</i> 0.05
	Observational Survey – Words	0.02	
	Observational Survey – Concepts about Print	-0.21	
Vadasy-97a	WJ-R – Word Attack	0.35	<i>Combined Mean</i> 0.47
	Dolch Word Recognition	0.31	
	Yopp-Singer Word Segmentation	0.74	
Vadasy-97b	WJ-R – Word Attack	0.21	<i>Combined Mean</i> 0.32
	Dolch Word Recognition	0.15	
	Yopp-Singer Word Segmentation	0.41	
	Bryant pseudoword	0.54	
	Pseudoword list	0.28	
Vadasy-00	WJ-R – Word Attack	1.21	<i>Combined Mean</i> 0.99
	Dolch Word Recognition	0.81	
	Yopp-Singer Word Segmentation	0.90	
	Bryant pseudoword	1.04	

読解力：読解力

調査	指標	<i>Unbiased d</i>
Allor-04.1	WJ-R Passage Comprehension	0.50
Allor-04.2	WJ-R Passage Comprehension	0.45
Baker-00	WRMT-R Word Comprehension	0.41
	WRMT-R Passage Comprehension	0.28
	<i>Combined Mean</i>	0.34
Cobb-01.2	GRAT-Reading Comprehension (GCOMP)	-0.22
Cobb-01.3	GRAT-Reading Comprehension (GCOMP)	-0.52
Mayfield-00	WRMT-R Passage Comprehension	0.53
Mehran-88	WJPEB Passage Comprehension	0.44
	CTBS Comprehension	0.09
	<i>Combined Mean</i>	0.27
Nielson-92	Stanford Achievement Test - Comprehension	0.28

読解力：オーラル・フルーエンシー

調査	指標	<i>Unbiased d</i>
Allor-04.1	Curriculum-based oral reading fluency	0.36
Allor-04.2	Curriculum-based oral reading fluency	0.30
Baker-00	Oral reading fluency – 1st grade passage Oral reading fluency – 2nd grade passage <i>Combined Mean</i>	0.45 0.47 0.46
Erion-94	Oral reading fluency	0.27
Morris-90.1	Basal passages	0.77
Morris-90.2	Basal passages	1.00
Rimm-Kaufman-99	Observational Survey – Reading Level	0.37
Vadasy-97a	Analytic Reading Inventory (fluency)	0.42
Vadasy-97b	Analytic Reading Inventory	0.15
Vadasy-00	Analytic Reading Inventory – Primary level Analytic Reading Inventory – 1st grade level <i>Combined Mean</i>	0.56 0.40 0.48
Weiss-89	Curriculum-based Measurement (CBM) Raw	-0.15

作文力

調査	指標	<i>Unbiased d</i>
Morris-90.1	Spelling (correct score) Spelling (qualitative score) <i>Combined Mean</i>	0.61 0.83 0.72
Morris-90.2	Spelling (correct score) Spelling (qualitative score) <i>Combined Mean</i>	0.83 0.72 0.77
Rimm-Kaufman-99	Observational Survey – Writing Observational Survey – Dictation <i>Combined Mean</i>	-0.01 0.18 0.08
Vadasy-97a	WRAT-R Spelling (Standard) Writing sample - Words written Writing sample - Words correctly spelled <i>Combined Mean</i>	0.78 0.44 0.37 0.53
Vadasy-97b	WRAT-R Spelling (Raw) Writing sample - Words written Writing sample - Words correctly spelled <i>Combined Mean</i>	0.20 0.26 0.18 0.21
Vadasy-00	WRAT-R Spelling Subtest (standard) Curriculum-based spelling measure Words correct (%) on writing measure <i>Combined Mean</i>	0.80 0.91 0.81 0.84

算数力

調査	指標	<i>Unbiased d</i>
Mahoney-86	Multiplication test (掛け算テスト)	0.69
McKinney-95	Stanford Achievement Test (8th ed.)	-0.14
Parham-94.1	Orleans-Hanna Algebra Prognosis Test	.09
Parham-94.2	Orleans-Hanna Algebra Prognosis Test	.96
Ritter-00	Stanford Achievement Test (9th ed.)	-0.12

12. 除外された調査： その特性と参考事項

表7： 除外された調査の特性の要約

	調査名／プログラムの簡単な説明	生徒	指導員	除外理由
1	Abbott, S. P., & Berninger, V. W. (1999). 処遇教育に遅すぎるといふことは決してない： 読解力障害を持つ4-7学年生に単語を覚えさせる。失語症の記録(Annals of Dyslexia), 49, 223-50.	Gr 4-7	学校付属の心理学専門医及び心理学専攻の大学院生	<ul style="list-style-type: none"> 指導員はボランティアではなかった。 2種類の処遇が比較された
2	Al-Hazza, T. C. (2003) America Reads の個別指導プログラムと指導員研修の都市部少数民族の落ちこぼれ寸前生の生活態度と学業成績に与える影響の調査(博士論文、オールドドミニオン大学、2002)。Dissertation Abstracts International(学位論文概要インターナショナル), 63, 10A.	Gr K-3	America Reads 大学生	<ul style="list-style-type: none"> 無作為実地試験にあらず(処遇校2校と対照校2校とのSESと学業成績に関する比較)
3	Allen, A., & Chavkin, N. F. (2004). 地域社会ボランティアによる個別指導が中学生の学力向上に役立つことを示す新たな証拠。School Community Journal, 14(2), 718.	中学生	America Reads 大学生	<ul style="list-style-type: none"> 無作為実地試験にあらず(個別指導の受講経験の多い生徒と少ない生徒との比較)
4	Berninger, V. W., Abbott, R. D., & Whitaker, D. (1995). 作文力障害児童用指導要領における低レベル能力と高レベル能力の統合。Learning Disability Quarterly(障害学習季刊誌), 18(4), 293-309.	小学生		<ul style="list-style-type: none"> 指導員はボランティアではなかった(2人の熟練臨床医)。
5	Bradley, K. L. (2002). 落ちこぼれ寸前の小学校4-5年生の読解力アップのための“一人でも良い、助けよう”プログラム(HOSTS)の効果(博士論文、オールドドミニオン大学、2001)。Dissertation Abstracts International, 63, 2A.	Gr 4-5	地域社会ボランティア (年齢まちまち)	<ul style="list-style-type: none"> 無作為実地試験にあらず(HOSTSの対象生39人のSESについての対照群との比較)
6	Burns, M. K., Senesac, B. V., & Symington, T. (2004). 読解力で落ちこぼれ寸前の子供の読解力向上努力におけるHOSTSプログラムの実効性。Reading Research & Instruction(読解力の研究と指導), 43(2), 87-104.	Gr K-5	地域社会ボランティア (年齢まちまち)	<ul style="list-style-type: none"> 無作為実地試験にあらず (プログラムを既に使用している6校を使用していない4校と比較)
7	Cobb, J. B. (2000). 指導員の専攻科目が指導員の指導能力に及ぼす影響： 落ちこぼれ寸前の4学年生用読み書き能力介入プログラム。 Journal of Reading Education(読解力教育ジャーナル), 25(3), 22-31.	Gr 4	教職実習生及び大学生運動選手	<ul style="list-style-type: none"> 対照群無し； 2人の異なるタイプの指導員による2つの処遇間の比較

	調査名／プログラムの簡単な説明	生徒	指導員	除外理由
8	Community Service Society of New York. (2000). The experience corps (経験部隊) : ニューヨーク南ブロンクスで好評の或世代間プログラム : Urban Agenda Issue Brief(都市政策問題報告) (ERIC Document Reproduction Services(ERIC書類複製サービス) No. ED476328)	Gr 1	退職者	<ul style="list-style-type: none"> 統計値の報告不備 ; 平均値のみ提供 (Meier and Invernizzi, 2001の論文に言及している様子)
9	Compton, G. L. (1992) The Reading Connection: 落ちこぼれ寸前の子供達向け教育サービスの提供法変更の上層部構想(博士論文, 西ミシガン大学, 1992)。 Dissertation Abstracts International, 53, 4A.	Gr 1	教員養成学校生	<ul style="list-style-type: none"> 統計値の報告不備 ; 記述統計を報告している唯一の表にある数値が実現不可能な値(自由度)
10	Rabiner, D. L., Malone, P. S., and the Conduct Problem Prevention Research Group (2004). 注意力欠陥障害のある及び無い子供に早期に読解力をつけるために行われる個別指導の影響。異常児の心理ジャーナル, 32(3), 273-284.	Gr 1	セミプロ	<ul style="list-style-type: none"> 指導員はボランティアでなく、ウォラック式教授法(Wallach and Wallach method) の訓練を40時間受けたセミプロであった。
11	Edwards, H. M. (2001). テキサス州ベクサー郡の落ちこぼれ寸前の小学生に実施された、選抜改善変数(selected improvement variables)に関する軍・学協力の個別指導・薫育イニシアティブの効果(博士論文, テキサス州A&M大学, 2000). Dissertation Abstracts International, 61, 11A.	Gr 3-5	テキサスの軍施設からの指導員、年齢 20-60	<ul style="list-style-type: none"> 無作為実地試験にあらず ; 分析用に個別指導の対象グループの下位集合(subset)を無作為に選び、個別指導の非対象集団から無作為に選ばれた下位集合と比較している。
12	Fantuzzo, J. W., Davis, G. Y., & Ginsburg, M. D. (1995). 生徒の自己概念と算数の成績に対する保護者の単独関与あるいは仲間個別指導との組合せによる関与の効果。教育心理学ジャーナル, 87(2), 272281.	Gr 4-5	仲間個別指導	<ul style="list-style-type: none"> ボランティアによる個別指導プログラムではなく、仲間個別指導プログラム。
13	Ferrier, M. H. (1994). Mentoring(薫育) : 落ちこぼれ寸前の小学生ならびに薫育の対象生のmentors (メンター)や teachers (教員)に与える効果。(博士論文, テキサス州A&M 大学, 1993). Dissertation Abstracts International, 54, 8A.	Gr 3-5	テキサス州ケリー空軍基地から派遣されたMentors(メンター)	<ul style="list-style-type: none"> 無作為実地試験にあらず ; 分析用に個別指導の対象グループの下位集合を無作為に選び、個別指導の非対象集団から無作為に選ばれた下位集合と比較している。
14	Fowler, M. C., Lindemann, L. M., Thacker-Gwaltney, S., & Invernizzi, M. (2002). マンツーマンの個別指導の2年目 : 小学校2年生の読解力不足児童に対する介入。CIERA レポート。ミシガン州アナーバー : 早期読解力成績向上センター(ERIC Document Reproduction Service(エリック文書複製サービス) No. ED468418).	Gr 2	地域社会ボランティア及び連邦労働学習大学院生(Federal Work Study Graduate Students)	<ul style="list-style-type: none"> 無作為実地試験にあらず ; 2年生の教員や校長が生徒を介入に割り当てる際に口出ししている(処遇の対象に選ばれたN=26人の児童は前テストの得点では他の児童の得点と似通っていたに拘わらず、より“介入の必要性大である”と判断されている。)

	調査名／プログラムの簡単な説明	生徒	指導員	除外理由
15	Invernizzi, M., Rosemary, C., & Juel, C., & Richards, H. C. (1997). 読解力不足生と地域社会ボランティア：3年間の展望。Scientific Studies of Reading (読解力の科学的調査), 1(3), 277-300.	Gr 1	地域社会ボランティア (年齢まちまち)	● 無作為実地試験にあらず(個別指導の受講経験の多い生徒と少ない生徒との比較)
16	Jason, L. A., Danner, K. E., & Kurasaki, K. S. (1993). 高リスクの編入生向け予防プログラムの1年間の追跡調査。情緒・行動障害ジャーナル, 1(4), 215-221.	Gr 3-5 (編入生)	大学生と保護者	● 関連性のあるアウトカム指標無し；読解力の点数があるのみ。
17	Jason, L. A., Johnson, J. H., & Danner, K. E. (1993). 包括的、予防的、保護者中心の高リスク編入生向き介入。福祉サービスにおける予防, 10(2), 27-37.	Gr 3-5 (編入生)	大学生と保護者	● 関連性のあるアウトカム指標無し；対応戦略のみ
18	Jason, L. A., Weine, A. M., & Johnson, J. H. (1993). 学校の変革プロジェクト：包括的予防的介入。情緒・行動障害ジャーナル, 1(1), 65-70.	Gr 3-5 (編入生)	大学生と保護者	● 統計値の報告不備；オンラインの論文に記述統計の記載無し；全体的読解力の指標に関する情報が必要。
19	Jason, L. A., Betts, D., & ; Johnson, J. (1989). 適応指導と個別指導の学校を中心とする予防プログラムの評価。Professional School Psychology(プロの学校心理学), 4(4), 273-284.	Gr 3-5 (編入生)	大学生と保護者	◆統計値の報告不備；記述統計の記載無し
20	Johnson, J. H., Jr. (1992). 低学年編入生の適応問題の実験的調査：パス解析を用い、クラスター分析に基づき編入生のサブタイプ分類を行っての編入生モデルの開発(博士論文、デポール大学, 1991). Dissertation Abstracts International, 52, 10B.	Gr 3-5 (編入生)	大学生と保護者	● 関連性のあるアウトカム指標無し；あるのは読解力と算数力の学校での成績といくつかの行動の指標
21	Juel, C. (1996). 読み書きの個別指導を効果的にするものは何? Reading Research Quarterly(読解力研究季刊誌), Vol 31(3), 268-289.	Gr 1-2	大学生-運動選手	● 無作為実地試験にあらず；同等でない比較グループとのQED(個別指導の対象グループを残りの生徒と比べている)
22	Lauren, L., & Allen, L. (1999). 早期読み書き介入プロジェクトの成功を予測する決め手となる要因。Reading Research Quarterly, 34(4), 404-424.	Gr 1-4	教職実習生	● 無作為実地試験にあらず；組立られた対照群とのQED(前テストの得点に合わせて)
23	Law, M., & Kratochwill, T. R. (1993). ペア読書：ある保護者個別指導プログラムの評価。School Psychology International, 14(2), 119-147.	Gr 2-4	ペア読書の訓練を受けた保護者	● 無作為実地試験にあらず；対照群無しの前・後比較

	調査名／プログラムの簡単な説明	生徒	指導員	除外理由
24	Leach, D. J., & Siddall, S. W. (1990). 読解力の指導への保護者参加：聞き取り、ペア読書、ポーズ、促し、ほめ言葉、及び直接指導法の比較。 <i>British Journal of Educational Psychology</i> (教育心理学英国ジャーナル), 60, 349-355.	Gr 1	保護者	<ul style="list-style-type: none"> • 米国で行われた調査でない。
25	Meier, J. D., & Invernizzi, M. (2001). ブロンクスの読書仲間： America Readsのあるモデルのテスト。 <i>Journal of Education for Students Placed at Risk</i> (落ちこぼれ寸前の生徒のための教育ジャーナル), 6(4), 319-333.	Gr 1	退職者	<ul style="list-style-type: none"> • 統計値の報告不備；記載されていたのは平均値のみ(及び報告されていた自由度に基づくと実施不可能なF-統計)。
26	Meyer, B. J., Middlemiss, W., & Theodorou, E. (2002). 5学年の児童を対象にインターネットで年輩の成人指導員がついたり、つかなかったりして実施された構造戦略指導(structure strategy instruction)の効果。 <i>Journal of Educational Psychology</i> , 94(3), 486-519.	Gr 5	退職者	<ul style="list-style-type: none"> • 対面個別指導プログラムの調査ではない(インターネット中心のプログラムである)。
27	Rembert, W. I. (1989). 年齢横断的個別指導と若年児童のロゴ・プログラミング環境における空間的問題の対応能力。博士論文、在グリーンズボロ北カロライナ大学		大学生を含む各種年齢層の人々	<ul style="list-style-type: none"> • 個別指導プログラムが1ヶ月の最小持続期間に満たなかった(プログラムは3週間しか続かなかった)。
28	Rogers, S. F. (1999). 読解力の実習課目の説明責任の変化：教職実習生は児童の進歩状況の責任を問われる。サウス・カロライナ州ヒルトンヘッド： College Reading Association(大学読書協会)の年次講座で発表された論文。(ERIC Document Reproduction Services No. ED447459)	Gr 2,5	教職実習生	<ul style="list-style-type: none"> • 無作為実地試験にあらず； 対照群無しの前・後比較
29	Ross, S. M., Morrison, G. R., Smith, L. J., & Cleveland, P. (1990). 落ちこぼれ寸前の小学生児童向け密着・突き放し交互個別指導の評価。 <i>Computers in Human Behavior</i> (人間行動におけるコンピュータ), 6(3), 247-259.	Gr 6	大学院生	<ul style="list-style-type: none"> • 対面個別指導プログラムの調査ではない(コンピュータ・ネットワークを利用した個別指導プログラムを電話会議を組み込んだプログラムと比較している)。 • 関連性のあるアウトカム指標無し；用法と態度を重視。
30	Taylor, J., & Cox, B. D. (1997). 小学校4年生による複雑な2段階数学用語問題のグループ・ベースの解法のMicrogenic analysis(マイクロジェニック分析)。 <i>Journal of the Learning Sciences</i> (学習の科学ジャーナル), 6(2), 183-226.	Gr 4	複数の教員と一人の心理学者	<ul style="list-style-type: none"> • 指導員はボランティアでなく、複数の教員と一人の心理学者であった。 t

	調査名／プログラムの簡単な説明	生徒	指導員	除外理由
31	Tucker, C. M., Herman, K. C., Reid, A. D., Keefer, N. L., & Vogel, D. L. (1998). 研究中心のモデル・パートナーシップ教育プログラム： 4年間の成績調査。Journal of Research and Development in Education(教育研究・開発ジャーナル), 32(1), 32-37.	Gr 3,9	大学生	<ul style="list-style-type: none"> • 無作為実地試験にあらず； 対照群にプログラムへの参加を申し込まなかった者が含まれている。 • 関連性のある成績無し(算数と読解力の点数のみ)。 • 3学年と9学年の生徒の成績を分離するすべ無し。
32	Vadasy, P. F. (2002). 有望な革新の持続可能性, 1998年11月1日-2002年10月31日： 最終報告、ワシントン州シアトル： ワシントン研究所 (ERIC Document Reproduction Services No. ED 474358)	Gr 1	地域社会	<ul style="list-style-type: none"> • プログラムの評価ではない；むしろ対照群無しの個別指導処遇の要約である。
33	Vadasy, P. F. (2001) 読解力問題を抱える2学年生向けの理解への道筋： 反復読書のマンツーマン指導対理解戦略指導(博士論文, ワシントン大学, 2000). Dissertation Abstracts International, 61, 11A.	Gr 2		(真の比較グループを備えた調査ではない；この調査は2種類の処遇を比較している。
34	Vinograd-Bausell, C. R., & Bausell, R. B. (1987). 単語の認識力の家庭内指導。Journal of Research & Development in Education, 20(3), 57-65.	Gr 1	保護者	(個別指導プログラムが最小継続期間1ヶ月に満たなかった(2週間続いただけ)。
35	Weine, A. M., Kurasaki, K. S., Jason, L. A., & Danner, K. E. (1993). 編入生向け予防的個別指導プログラムの評価。Child Study Journal(児童研究ジャーナル), 23(2), 135-152.	Gr 3-5 (編入生)	大学生と保護者	<ul style="list-style-type: none"> • 統計値の報告不備； 記述統計の記載無し。

13. 全体的参考文献

- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A metaanalysis of findings. *American Educational Research Journal*, 19, 237-248.
- Devin-Sheehan, L., Feldman, R. S., & Allen, V. L. (1976). Research on children tutoring children: A critical review. *Review of Educational Research*, 46, 355-385.
- Elbaum, B., Vaughn, S., Hughes, M.T., & Moody, S.W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A metaanalysis of the intervention research. *Journal of Educational Psychology*, 92(4), 605-619.
- Ellson, D.G. (1976). Tutoring. In N. Gage (Ed.). *The psychology of teaching methods*. Chicago: University of Chicago Press.
- Fashola, O.S. (2001). *Building effective afterschool programs*. Thousand Oaks, CA: Corwin Press.
- Fitz-Gibbon, C.T. (1977). *An analysis of the literature of cross-age tutoring*, Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED148807).
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Goodlad, S. (Ed.). (1995). *Students as tutors and mentors*. London, PA: Kogan Page.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Michael, B. (1990). *Volunteers in public schools*. Washington, DC: National Academy Press.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- National Research Council. (1998). *Preventing reading difficulties found in children*. Washington, DC: National Academy Press. Retrieved June 21, 2006, from <http://www.readbygrade3.com/prevread.htm#contents>
- Rekrut, M. D. (1994). Peer and cross-age tutoring: The lessons of research. *Journal of Reading*, 37(5), 356-362.
- Rosenshine, B., & Furst, N. (1969). *The effects of tutoring upon pupil achievement: A research review*. Washington, DC: Office of Education. (ERIC Document Reproduction Service No. ED 064 462).
- Rothstein, H. R., Turner, H. M., & Lavenberg, J. G. (2004). *The Campbell Collaboration information retrieval policy brief*. Philadelphia, PA: The Campbell Collaboration.
- Shanahan, T. (1998). On the effectiveness and limitations of tutoring in reading. In P. D. Pearson & A. Iran-Nejad (Eds.). *Review of Research in Education* (pp. 217-234). Washington, DC: American Educational Research Association.
- Topping, K. J., & Hill, S. V. (1995). University and college students as tutors for schoolchildren: A typology and review of evaluation research. In S. Goodlad (Ed.), *Students as tutors and mentors* (pp. 13-31). London, PA: Kogan Page.
- U.S. Department of Education. (1996). *Examples of tutoring programs*. Washington, DC: America Reads Challenge (<http://www.ed.gov/inits/americanreads/readexam.htm>, Updated August 15, 1996).
- U.S. Department of Education. (1997). *Read with me: A guide for student volunteers starting early childhood literacy programs*. Washington, DC: Office of Educational Research and Improvement.
- Wasik, B. A. (1998). Volunteer tutoring programs in reading: A review. *Reading Research Quarterly*, 33,

266-292.

Wasik, B. A., & Slavin, R. A. (1993). Preventing early reading failure with one-to-one tutoring: A review of five great programs. *Reading Research Quarterly*, 28(2), 179-200.

White House, Office of the Press Secretary. (1997). *President Clinton's America Reads Work-Study Challenge*. Washington, DC: Author.

Wilson, D., & Lipsey, M. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods*, 6(4), 413-429.

14. 共同レビューワの連絡先情報

Dr. George Denny
Professor, Educational Statistics
University of Arkansas
234 Graduate Education Building
Fayetteville, AR 72701
Phone: (479) 575-7320
Fax: (479) 575-2492
Email: gdenny@uark.edu

Ginger Albin
Graduate Student, Higher Education
University of Arkansas
234 Graduate Education Building
Fayetteville, AR 72701
Phone: (918) 809-2239
Fax: (479) 575-2492
Email: galbin@uark.edu

Joshua Barnett
Graduate Assistant, Office for Education Policy
University of Arkansas
201 Graduate Education Building
Fayetteville, AR 72701
Phone: (479) 575-3172
Fax: (479) 575-3196
Email: jhb05@uark.edu

Virginia Blankenship
Research Associate, Office for Education Policy
University of Arkansas
201 Graduate Education Building
Fayetteville, AR 72701
Phone: (479) 575-3773
Fax: (479) 575-3196
Email: vhudson@uark.edu

15. 目次

表紙.....	1
概要.....	3
要約.....	3
背景.....	3
目標.....	3
検索方法.....	3
選択規準.....	3
データの収集と分析.....	4
主な結果.....	4
レビューワの結論.....	4
1. 背景.....	6
2. 目標.....	8
3. レビューの方法.....	8
調査のレビューの対象としての採用規準.....	8
調査の種類.....	8
参加者の種類.....	8
介入の仕方.....	9
アウトカム指標の種類.....	9
調査の特定に利用した検索方法.....	10
試験の選択.....	12
調査方法の質の評価.....	14
データの管理と抽出.....	14
データの総合.....	15
等質性分析.....	15
感度分析.....	16
下位グループ分析.....	16
出版バイアス.....	16
調査データの記載不備.....	16
下位グループとモデレーターの事後分析.....	16
4. 結果.....	17
対象に選ばれた調査の説明.....	17
全アウトカムの総合効果量.....	19
全体的読解力.....	20
文字・単語読解力.....	20
読解力.....	20
読解力オーラル・フルーエンシー.....	21
作文力.....	21
全体的算数力.....	21
読解力アウトカムに関する調査の下位グループに対する影響分析.....	22
5. 論考と結論.....	23

6. 本レビューの更新計画	25
7. 謝辞	25
8. 潜在的利益相反	25
9. 本レビューの対象になった調査：特性と参考事項.....	26
参考文献リスト - 対象となった調査(21 論文).....	35
10. 図：すべての重要アウトカムの森林プロット.....	37
11. 追加表	47
12. 除外された調査：その特性と参考事項.....	51
13. 全体的参考文献	56
14. 共同レビューワの連絡先情報.....	58
15. 目次	59