*Review of Educational Research*
*Winter 1998, Vol. 68, No. 4, pp. 450–469*

# Effects of Instruction in Deriving Word Meaning from Context: A Meta-Analysis

**R.G. Fukkink**
**K. de Glopper**
*University of Amsterdam*

*A meta-analysis of 21 instructional treatments aimed at enhancing the skill of deliberately deriving word meaning from context during reading shows a medium effect size of 0.43 standard deviation units ($p < .000$). An exploratory multilevel regression analysis shows that clue instruction appears to be more effective than other instruction types or just practice ($\beta = 0.40$). Effect size correlates negatively with class size ($\beta = -.03$). Implications for instruction and future research are discussed. Future studies should investigate the effect of instruction on both the skill of deriving word meaning from context and incidental word learning to evaluate its contribution to vocabulary growth.*

Educational researchers in the field of vocabulary learning emphasize different instructional activities to expand the reading vocabulary of students. In their review of vocabulary instruction, Beck and McKeown (1991) identified four prominent positions. The first position promotes direct instruction in word meanings (e.g., Zechmeister, Chronis, Cull, d'Anna, & Healy, 1995). Proponents of the second position consider wide reading as an important source for vocabulary learning and assume that readers learn words incidentally during normal reading by deriving word meaning from context (e.g., Nagy, Anderson, & Herman, 1987). Advocates of the third position oppositely stress the intrinsic complexity of deciphering the meaning of unknown words and encourage students to look up their meanings in a dictionary (e.g., Schatz & Baldwin, 1986). The fourth position is a synthesis of the second and third position. Some researchers acknowledge the potential value of deriving meaning from context, but are aware of the complexity of this task, and therefore advocate instruction to enhance this skill.

A considerable amount of research on vocabulary instruction has been devoted to direct instruction of word meanings. Meta-analyses have shown that vocabulary instruction produces strong effects on vocabulary and small effects on reading comprehension for both average (Stahl & Fairbanks, 1986) and poor readers (Marmolejo, 1990). However, the limitations of this type of instruction are obvious. According to Stahl and Fairbanks, a vocabulary teaching program typically teaches 10 to 12 words per week or about 400 per year of which perhaps 75 percent are learned. The number of words to be learned by students, however, is much too high to teach all of them. Given the size of the task, no vocabulary teaching program alone can produce the vocabulary growth that is necessary to

become a proficient reader, and incidental word learning should therefore be promoted (Nagy, Herman,& Anderson, 1987).

The hypothesis that readers learn words incidentally during normal reading by deriving word meaning from context is investigated in a number of studies, starting with Jenkins, Stein, and Wysocki (1984). Research in this field has demonstrated that incidental word learning, which involves both the derivation of word meanings and memorization, is a relatively slow, incremental process which leads only to a strong cumulative effect across many years of wide reading because of the large number of unknown words that students encounter during reading (Shefelbine, 1990). Researchers do not agree on the exact contribution of incidental word learning to vocabulary growth. Estimates have been proposed from a low 160 words (Carver, 1994), through 800-1,200 words (Nagy, Anderson, & Herman, 1987) to 600-2,000 words each year (Shu, Anderson, & Zhang, 1995) for younger readers. Beck and McKeown (1991) estimate, based on a review of several studies, that children learn 2,700-3,000 words per year. This substantial variation is the consequence of differences in the estimates of the word learning rate, the amount of reading, the proportion of unknown words younger readers typically encounter in their reading materials. The estimate of the word learning rate is also influenced by the definition of what counts as a word (D'Anna, Zechmeister, & Hall, 1991).

Instruction in deriving word meaning from context to increase vocabulary is tightly connected to the incidental word learning hypothesis. Because students encounter a large number of words, even a small improvement of the ability to infer the meaning of unknown words would result in a sizable number of words learned. Deriving word meaning from context has therefore "a sound and persuasive rationale", according to Jenkins, Matlock, and Slocum (1989, p. 218), and many other authors have also acknowledged the potential value of instruction in deriving word meaning from context. Another argument for the importance of instruction in deriving word meanings is that, regardless of any impact on incidental word learning, students need strategies for coping with unfamiliar words encountered while reading.

Studies which focused at the process of deriving word meaning, have made clear that students experience many problems trying to decipher the meaning of unknown words (Van Daalen-Kapteijns, Schouten-van Parreren, & De Glopper, 1993). These problems can be attributed to the text or the reader. The context does not clarify the full meaning of unknown words (Beck, McKeown, & McCaslin, 1983; Schatz & Baldwin, 1986), and, second, readers have limited ability and experience in intèntionally deriving meaning from context (McKeown, 1985; Shefelbine, 1990; Van Daalen-Kapteijns & Elshout-Mohr, 1981). Further evidence for the complexity of the word meaning derivation task stems from individual word learning studies that employed a "context-only" method. In these studies, students must learn the meaning of words by reading contrived texts that contain the target words, and no further provisions are undertaken to teach their meanings. In contrast with direct instruction of word meanings, the "context-only" method shows rather weak results. Stahl and Fairbanks (1986) report mixed results for the "context-only" studies. The meta-analysis of Marmolejo (1990) reports a small, nonsignificant result for these studies with poor readers ($d = .11$).

The focus of this article is the effect of instruction on the skill of deriving word meaning from context during reading. Does instruction improve the skill in intentional word learning? Comments in the literature on the effects of instruction on this skill are predominantly pessimistic. Graves (1986) concludes from his summary of research that

> teaching students to use context is difficult. In fact, there is no report that presents a thorough and convincing case that students can be taught to better use context to unlock the meanings of novel words encountered during normal reading" (p. 73).

Beck and McKeown (1991) conclude in their review that "at best, researchers have found admittedly small gains" (p. 803). Carnine, Kameenui, and Coyle (1984) stated that "experimental methodology and the instructional procedures used in the intervention research have been diverse and have not resulted in identifying a specific set of instructional strategies for teaching students how to use context clues in understanding the meaning of an unfamiliar word" (p. 196). The recent review of Kuhn and Stahl (1998) is in line with these former studies. They state that if the reviewed studies 'represent where the field is now, then we cannot recommend instruction in context clues' (p. 135). Kuhn and Stahl conclude cautiously that practice may be equally effective as instruction. They tend to favor the interpretation that 'it is likely that students benefit as much from practice in deriving words from context as they would from instruction in either a specific set of strategies or a list of clues' (p. 129). Their interpretation is motivated by their finding that the studies that included a practice-only group did not find statistically significant differences between the experimental treatment and the practice-only condition.

The abovementioned reviews have two limitations. First, some relevant studies were not included in the reviews. Secondly, reviews were of a narrative style, and a vote-counting method was used on a few studies with small sample sizes. The statistical power of the reviewed studies is small, and statistical significance of the results is therefore not the best method to evaluate findings. This article summarizes by the method of meta-analysis the results of interventions involving the skill of deliberately deriving word meaning from context during reading. In an explorative analysis, the study outcomes are subsequently interrelated with the methodological and instructional characteristics of the studies using multilevel regression analysis (Bryk & Raudenbusch, 1992).

## Method

### *Sample of Studies*

Studies were identified from a computer search of the databases of ERIC (1965 - June 1997), Linguistics and Language Behavior Abstracts (1973 - July 1997), Dissertation Abstracts (1861- August 1997), and PsycLit (1974 - September 1997). These databases were searched with the profile '[read* and vocabulary]' and '[context* or infer* near meaning or deriv* near meaning]'. The keywords used were derived from key publications, cited in the reviews of Graves (1986) and Beck and McKeown (1991). This search was followed by cross-referencing of located studies. Finally, a search was conducted on relevant journals to locate recent studies on the topic of interest.

TABLE 1
*Reliability of coding by study characteristic: κ/$r_I$ and agreement rate (AR)*

| | κ[a] | $r_I$[b] | AR (%)[c] |
|---|---|---|---|
| Design of study | 1 | - | 100 |
| Random assignment | 1 | - | 100 |
| Adjusted means | .75 | - | 93 |
| Type of test | 1 | - | 100 |
| Dependency of items | .67 | - | 79 |
| Internal consistency | - | .99 | 75 |
| Class size | - | .98 | 86 |
| Age | - | .89 | 91 |
| Type of instruction | .92 | - | 91 |
| Instructional time | - | .99 | 71 |
| Control group | .53 | - | 64 |

[a] = generalized Cohen's Kappa for nominal variables
[b] = generalized intraclass correlation (design 2) (Orwin, 1994) for continuous variables
[c] = percent of observations agreed upon by the three coders for nominal and continuous variables

In order to be included in the meta-analysis, a study had to meet the following eligibility criteria. First, the treatment must aim specifically at enhancing the skill of deliberately deriving word meaning from context during reading. Second, this skill must be adequately measured at the posttest. Studies were not included if the posttest contained target words that had been the subject of explicit instruction or if the test administered was a broad measure of many abilities from which scores for deriving word meaning could not be deduced (e.g., Farr, 1987) or if a metalinguistic test was administered (e.g., Carr & Mazur-Stewart, 1988; see also the comments on 'linguistic versus metalinguistic skill' by Kuhn & Stahl, 1998) or if the target skill was not measured at all (e.g. Askov & Kamm, 1976; Gifford, 1993). Third, only studies with a control group design were included, leaving out Goerss (1995) in which a no control group design is applied. In addition, the control group should receive no instruction aimed at enhancing the skill of interest. Friedland (1992) compared the differential effects of two different treatments without including a control group, which may have led to reduced effect sizes that are not comparable to those from other studies. Finally, the studies of Hafner (1965); Patberg, Graves, and Stibbe (1984); Sternberg (1987); and Wheatly, Muller, and Miller (1993) were not included because their reports did not provide the necessary statistics.

Not all studies could be obtained. The studies of Peterson (1942; unpublished doctoral dissertation), Butler (1943; unpublished master's thesis), and Jensen (1947; unpublished master's thesis), discussed by Guarino (1960), could not be obtained. The paper of Patberg and Stibbe (1985; cited in Beck & McKeown, 1991) and Frye (1975) could not be acquired either.

Twelve studies that investigated the effects of instruction on the skill in deriving word meaning from context were finally included in the meta-analysis. Guarino (1960) and Schwartz and Raphael (1985) concern two different experiments and six studies are multiple-treatment studies investigating the effects of different experimental treatments in one study. The number of experimental treatments is 22. The treatment is the unit of analysis for determining effect sizes and subsequent statistical analysis (see Table 2).

### *Computation of Effect Sizes*

The effect size calculated is d, based on the pooled sample standard deviation, and adjusted for bias due to small samples (Hedges & Olkin, 1985). All effect sizes and standard errors were determined from the reported means and standard deviations using the program META from Schwarzer (1989).[1] Carnine et al. (1984) report in their study scores for five related measures. The five effect sizes were collapsed in one unweighted average effect size. Jenkins et al. (1989) report scores for four small tests, which were also summarized in one unweighted average effect size. For the study of Herman and Weaver (1988), effect sizes were calculated for the experimental 'After strategy instruction group', using the 'Before strategy group' as control group. The 12 studies yielded 22 experimental treatments with 22 effect sizes for deriving word meaning.

### *Coding*

Treatments were classified by characteristics of methodology, educational setting, and instruction. The following methodological characteristics were coded: design of study (pretest-posttest design with a parallel or identical test, or other), assignment to treatments (random assignment or blocking at student level or other), adjustment of means (correction for initial differences by covariance analysis, or not), type of dependent measure (cloze test, multiple choice test, or definition task), dependency of test items (several items from a test stem from one text, or not), internal consistency of the test, and initial differences at the pretest between the control and experimental group. Reliability figures were estimated for the studies of Buikema and Graves (1993), Carnine et al. (1984), Herman and Weaver (1988), Jenkins et al. (1989), and Schwartz and Raphael (1985) by inserting into the Spearman-Brown formula an estimate of the mean inter-item correlation, based on studies that report Cronbach's α and the number of items (.134). This estimate is derived from the studies that reported both the internal consistency and the number of items of the test. The initial difference between the control and experimental group was also computed to analyze possible covariation between initial differences and effect sizes at the posttest. This allows for a test of bias due to a possible lead of the experimental group at the start of the study. The initial difference is expressed as Hedges's *d*. The initial differences relate to pretests measuring derivation skill for pretest-posttest designs and proxy pretests for other designs (reading comprehension tests, vocabulary scores) (see Table 2). Carnine et al. (1984) do not report statistics for the pretest, but because student assignment is random, initial differences can be assumed to be small, and d is set to null. Jenkins et al. (1989) report nonsignificant differences at the pretest ($F$ [5,107] = 1.34 and 1.48 for the vocabulary and reading pretest respectively), and here *d* is set to null too.

With respect to educational setting, we coded: age (in years), class size, and

TABLE 2
*Summary of features and effect sizes of the studies included in the meta-analysis*

| Study | $N_e$+$N_c$ | PPD | A | Mean | Test | α | Pre | Age | CS | Dur. | Type | d | se |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Guarino (1960): Syracuse study | 85+82 | 1 | NR | NA | MC | .83 | .34 | 15 | 28.3 | 360 | clue | .60 | .16 |
| Guarino (1960): Michigan study | 153+148 | 1 | NR | NA | MC | .79 | .02 | 15 | 25.5 | 495 | clue | .32 | .12 |
| Cox (1974): Cloze 1-group | 25+22 | 0 | R | NA | MC | .84 | -.06 | 9 | 25 | 720 | cloze | -.12 | .29 |
| Cox (1974): Cloze 2-group | 24+22 | 0 | R | NA | MC | .84 | -.07 | 9 | 24 | 720 | cloze | .21 | .30 |
| Bissell (1982): Cloze | 30+28 | 1 | R | NA | MC | .38 | .00 | 18.5 | 23.5 | 238 | cloze | .00 | .26 |
| Bissell (1982): Forced Cloze | 29+28 | 1 | R | NA | MC | .38 | .00 | 18.5 | 23.5 | 238 | cloze | .23 | .27 |
| Sampson et al. (1982) | 46+46 | 1 | R | A | C | .85 | - | 8 | 7 | 662 | cloze | .73 | .22 |
| Carnine et al. (1984): rule | 12+12 | 0 | R | NA | MC | .55 | .00 | 10 | 1 | 90 | strategy | .82 | .43 |
| Carnine et al. (1984) practice | 12+12 | 0 | R | NA | MC | .55 | .00 | 10 | 1 | 90 | practice | .73 | .42 |
| Schwartz & Raphael (1985): Exp. I | 8+ 8 | 0 | R | NA | def | .55 | .00 | 9 | 8 | 160 | definition | 5.45 | 1.19 |
| Schwartz & Raphael (1985): practice | 14+28 | 1 | R | NA | def | .61 | -.04 | 10 | 14 | 160 | practice | .35 | .11 |
| Schwartz & Raphael (1985): training | 14+28 | 1 | R | NA | def | .61 | .59 | 10 | 14 | 160 | definition | 1.22 | .12 |
| Herman & Weaver (1988) | 14+16 | 0 | NR | NA | def | .13 | .15 | 13 | 14 | 325 | strategy | .57 | .37 |
| Kranzer (1988): textbook context | 19+19 | 0 | NR | NA | def | .66 | .24 | 13 | 23 | 155 | strategy | .50 | .33 |
| Kranzer (1988): enriched context | 21+19 | 0 | NR | NA | def | .66 | .01 | 13 | 23 | 155 | strategy | -.40 | .32 |
| Kranzer (1988): definition+context | 21+19 | 0 | NR | NA | def | .66 | .20 | 13 | 23 | 155 | strategy | .43 | .32 |
| Jenkins et al. (1989): low | 22+22 | 0 | NR | NA | def | .61 | .00 | 10 | 23 | 135 | strategy | .03 | .30 |
| Jenkins et al. (1989): medium | 22+22 | 0 | NR | NA | def | .61 | .00 | 10 | 23 | 165 | strategy | .48 | .31 |
| Jenkins et al. (1989): high | 22+22 | 0 | NR | NA | def | .61 | .00 | 10 | 23 | 300 | strategy | .71 | .31 |
| Buikema & Graves (1993) | 19+19 | 0 | NR | NA | MC | .61 | .28 | 12.5 | 19 | 250 | clue | .80 | .34 |
| De Glopper et al. (1997) | 29+30 | 1 | R | A | def | .78 | - | 11 | 8 | 480 | strategy | .15 | .26 |
| Tomesen & Aarnoutse (1998) | 16+15 | 0 | NR | NA | def | .84 | -.21 | 9 | 4 | 450 | clue | 1.53 | .41 |

*Note.* $N_e$ + $N_c$ = size of the experimental group + size of the control group; PPD: 1 = pretest posttest design, 0 = other design; Assignment: R = random, NR = not random; Means: A = mean are adjusted by covariance analysis; NA = means are not adjusted; Test: MC= multiple-choice, cloze= cloze test; def= definition task; Test: mc = multiple choice, c =cloze test, def = definition task; α = Cronbach's alpha of the posttest (estimated values in italics); Pre: difference at the pretest in standard deviation units between experimental and control group (estimated values in italics); Age = average age of students; CS = class size; Dur.: duration of the instruction in minutes; Type: type of instruction; d = effect size d; se = standard error of the effect size amount of instruction (in minutes). The sample of students in Bissell (1982) is formed by different age groups (62% freshmen, 26% sophomores, 9% juniors, rest unknown), and age is set at 18.5. Class size was determined directly from the report or indirectly by dividing the size of the experimental groups by the number of experimental treatments or by dividing the number in the experimental group by the number of classes. For the practice-only treatment in the study of Carnine et al. (1984), class size is determined to be 1. The amount of instructional time is determined directly from the reports in most cases. Carnine et al. do not report the exact amount of instructional time but report the number of lessons instead, namely three. For the treatments of this study, instructional time is estimated to be 90 minutes. No value could be estimated for Herman and Weaver (1988), and here the mean of the other treatments is inserted. The time of instruction might be of interest because it has been suggested that the instructional time that is normally devoted to this skill is probably too brief (Goerss,1995; Graves, 1986).

The instructional characteristics concern the type of instruction for the experimental group and the activities of the control group. Five types of experimental instruction are distinguished: context clue, cloze, strategy, definition, and practice-only instruction. In context clue studies, instruction and practice center on one or more context clue types. Students are taught to recognize and use certain context clues to elicit the meaning of an unfamiliar word. In some studies, clue instruction is incorporated in a generic strategy that emphasizes the recognition of the instructed clues[2] This type of instruction is closely aligned to studies in which a classification of context clues is proposed, and can therefore be labeled as text-oriented (see Rankin & Overholser, 1969 and Boettcher, 1980 for an overview; Humes, 1978; Sternberg, Powell & Kaye, 1983). Starting with McCullough as early as 1943 (as cited in Guarino, 1960), researchers suggest that different types of clues can be classified for teaching purposes. Context clue instruction is based on the assumption that clues exist that reveal the meaning of words in text, and readers should become aware of these clues to use them effectively. It may be possible that, by starting with some specific clues, clue instruction makes younger readers aware that context in general can serve as an aid, and therefore effects may transfer to texts without specific clues.

Cloze instruction aims to increase sensitivity to context by using cloze tests. These studies are "based on the assumption that by going through the task of completing cloze units, a reader will gain insights into the process of using context, recognizing the interrelationships of language, and consequently improving comprehension skills" (Jongsma, 1971; in Cox, 1974, p. 3).

In the strategy studies, instruction focuses at developing a general strategy to infer word meaning from context without explicit reference to clue types. This type of instruction is inspired by studies in which the process of deriving meaning from context is investigated, and can therefore be labeled as reader-based (see McKeown, 1985; Van Daalen-Kapteijns & Elshout-Mohr, 1981; Van Daalen-Kapteijns, Schouten-van Parreren, & De Glopper, 1997; Werner & Kaplan, 1952). This instruction is based on the rationale that readers become more capable of exploiting the context by providing them with a systematic approach.

Instruction in the definitional approach is designed to help students develop a general schema to conceptualize a definition. 'Students have only a vague concept of what constitutes a definition', and 'their acquisition of each new vocabulary item will be confounded by difficulties of selecting appropriate strategies, monitoring performance, and evaluating their attempts at definition', as Schwartz and Raphael (1985) state. The provided concept of definition 'becomes the basis for students to select and organize both background knowledge and context clues to construct the meaning of new vocabulary,' according to Schwartz and Raphael (p. 116).

In the practice-only type of instruction, students practice exercises without receiving further instruction. This kind of instruction was applied in one of the experimental treatments of the study of Carnine et al. (1984).

In classifying the studies, we also coded whether the control group receives no instruction or follows regular reading or language arts classes.

### *Interrater reliability*

All treatments have been coded independently by three coders. Generalized Cohen's kappa (Conger, 1980; Light, 1971 ) is computed for nominal categories, $r_I$(design 2) (Orwin, 1994) for continuous categories. Rate of agreement is given for both nominal and continuous categories. The results are summarized in Table 1.

A majority rule directed decisions in coding of both nominal and continuous scores. The characteristics 'activities of the control group' and 'dependency of items' could not be obtained from each report and reliably coded, and both were therefore dropped from further analysis.

## Description of the Studies Included in the Meta-Analysis

The methodological characteristics of the studies show variation (see Table 2). An untreated control group design with pretest and posttest is applied in four of 12 studies only. In the other studies, a proxy pretest design (Cook & Campbell, 1979) is used with reading comprehension or vocabulary scores as pretests. Unfortunately, random assignment of students to conditions is applied in only five studies. Only 10 out of the 19 treatments show statistically significant effects. However, with the exception of Guarino (1960), sample size is small and statistical power low for demonstration of moderate effect sizes.

Students to whom the instruction is given range from middle-grade to 10th grade with the exception of Bissell (1982), in which older students of 17 to 30 years old are instructed. The variation in class size is relatively large. The treatments can be divided into those with classes of normal size (> 20) and those in which more intensive instruction is delivered to small groups (4-8 students or even 1 to 1 tutoring). The amount of instruction is usually small (the average amount is 5.5 hours). Although the amount of instruction differs between treatments, it is not likely that an effect of instructional time will be found because of a restriction of range.

Clue instruction is employed in three treatments. In the Syracuse experiment of Guarino (1960), the definition, synonym, contrast, experience, and illustration clues were practiced. The latter clue was substituted for summary clues by Guarino in the Michigan experiment. The illustration clue is a category of the clue classification of Artley (1943; as cited in Guarino), whereas the other clues and their labels stem from the classification of McCullough (1943; as cited in Guarino). Buikema and Graves (1993) taught their pupils clues related to the action, the purpose, and the sensory aspects of a word, derived from the Sternberg and Powell

(1983) classification. Tomesen and Aarnoutse (1998) confined instruction to illustration, synonym, antonym clues, and general clues, excluding the knowledge of the world, logical consequence, and association clues of her classification.

The cloze procedure is employed as a teaching technique in five treatments. These treatments are all characterized by a relatively small amount of teacher instruction and an emphasis on practice. The cloze tests in these treatments are modified versions of the usual cloze test format in which words are deleted at fixed intervals (e.g., each ninth word). In the cloze practices used by Sampson et al. (1982), key words were deleted only. The cloze tests in the Forced Cloze treatment of Bissell (1982) combined a cloze test with a multiple choice test format. Students in this treatment could choose from two options filling in each cloze blank. It was hypothesized that this kind of practice

> would not only focus the uncertainty of the reader (by fostering selective attention to critical features in the sentence or passage) and facilitate self-monitoring (by allowing him to check the adequacy of his response as he reads), but that it would exert control over his future behavior on the task (p. 7).

Strategy instruction is most frequent among the treatments included in the meta-analysis. Carnine et al. (1984) is the study where this type of instruction in the derivation of word meaning was first applied. Their strategy involved a rule, like "When there's a hard word in a sentence, look for other words in the story that tell you more about that word" (p. 197). Students were also told that the unknown word gave information about a character in the story, or what and how something is done. Instruction in Herman and Weaver (1988) involved students to "'look in' at morphemes within a word" and "'to look around' at the flow of events and mood in the part of the story in which the word appeared" (p. 3). Both strategies were modeled by the teacher. Kranzer (1988) taught students a four-step strategy, summarized by the acronym 'SCAR': substitute, check the fit, accept the substitution, or rethink, if necessary. This strategy was adapted from Jenkins et al. (1989). In their study, a 'SCANR procedure' was taught that involved five steps: substitute a word or expression for the unknown word; check the context for clues that support your idea; ask if substitution fits all context clues; need a new idea?; and revise your idea to fit the context (Kranzer combined these two last steps into one step, 'rethink, if necessary'). Strategies in both studies were modeled by the teacher. The most elaborate strategy was taught in the study of Van Daalen-Kapteijns, Schouten-van Parreren, and De Glopper (1997). Instruction consisted of three strategies. First, students should apply a 'brake tactic' on encountering an unknown word. This tactic is followed by a 'track tactic', which consists of four consecutive steps. First, students are taught to make a 'substitution sentence'. On the basis of this sentence, students ask Wh-questions (where, why, who., etc.). From the information thus gathered, one should find a synonym or, if not available, a definition containing general and specific characteristics must be constructed. The fourth step is to check wether this synonym or definition fits the context. As a final strategy, students 'zoom in' on the word and check whether this yields information that confirms the word meaning derived so far.

The strategies in these studies differ on a number of aspects. The strategies can be divided in those that involve external clues only, or both external and internal clues (Herman & Weaver, 1988; Van Daalen-Kapteijns et al., 1997b). The instruction differs also in the number of strategies taught and the number of steps involved in them. The strategy in Carnine et al. (1984) is relatively simple. The strategies in Kranzer (1988) and Jenkins et al. (1989) involve a four- and five-steps procedure. The most complex procedure is applied in Van Daalen-Kapteijns et al. (1997b), where three strategies are taught, involving six steps all together.

TABLE 3
*Results of the random effects model*

| Fixed effect | Coefficient | se | t | p | 95-CI |
|---|---|---|---|---|---|
| Δ/grand mean | .43 | .084 | 5.15 | .000 | .25-.62 |
| Random effect | Variance component | df | $P^2$ | p | |
| | .061 | 20 | 36.90 | .012 | |

## Results of the Meta-Analysis

The effects of the treatments were analyzed using a random effects model. A random effects model is chosen for conceptual and empirical reasons. Because the studies involved in the meta-analysis differ on many characteristics, a random effects model is considered to be more appropriate for a priori reasons. The large variation in effect sizes that can not be explained by the fixed part of the unconditional model (see Table 3) is a further justification of this choice (Cooper & Hedges, 1994).

Preliminary examination of the effect sizes showed one statistical outlier, defined as values that are 1.5 interquartile range from Tukey's hinges (Tukey, 1977). The effect size $d = 5.45$ of Experiment I in Schwartz and Raphael (1985) is an extreme score, which is the result of a serious floor effect in the control group (the mean score of the control group is 0.4 with a standard deviation of 0.52, whereas the experimental group scored 12.9 with a standard deviation of 3.02). This treatment is therefore excluded from further statistical analysis.

The meta-analysis shows a significant positive effect for instruction in the skill of deriving word meaning from context. The generalized effect size Δ of 0.43 for skill in deriving is close to a 'medium' effect as defined by Cohen (1988) and Lipsey (1990). The results of the treatments are heterogeneous, as indicated by the $\chi^2$ value of the *Q*-test for homogeneity. Sampling error accounts for 50.9 percent of the variance. After correction for attenuation (Hunter & Schmidt, 1990), the generalized effect size Δ increases to 0.57 (*se* = .12; *p* = .000) with a 95%-confidence interval that ranges from .33 to .82.

### *Multilevel Regression Analysis of the Variation in the Effects*

Because the treatments show substantial variation in their outcomes, an exploratory analysis is undertaken to explain the heterogeneity of results, using multilevel regression analysis (Bryk & Raudenbusch, 1992). The first level is the treatment level with as corresponding model:

TABLE 4
*Results of the effect size analysis*

| Fixed effect | Coefficient | se | t | p | Variance explained |
|---|---|---|---|---|---|
| intercept | .882 | .187 | 4.71 | .000 | |
| class size | -.030 | .009 | - 3.13 | .006 | 24.4% |
| clue instruction | .400 | .165 | 2.42 | .027 | 58.8% |
| Random effect | Variance component | df | $\chi^2$ | p | |
| | .025 | 18 | 23.58 | .17 | |

$$d_j = \delta_j + e_j$$

where the effect size $d_j$ of the study j is an estimate of j with a sampling error $e_j$. At the second level, variation in outcomes between treatments is described. The true effect size in the Level-2 model depends on treatment characteristics and a Level-2 random error:

$$\delta_j = \gamma_0 + \gamma_1 W_{1j} + \gamma_2 W_{2j} + \ldots + \gamma_S W_{Sj} + u_j$$

where $W_{1j}$, ... $W_{Sj}$ are treatment characteristics of each study j, $\gamma_0$, ... $\gamma_S$ are their regression coefficients, and $u_j$ is a Level-2 random error.

The predictors are divided into a methodological subset (design of study, assignment, adjusted means, type of test, and internal consistency), an educational setting subset (class size, instructional time, and age), and an instructional subset that consists of the different types of instruction (see Table 2). Methodological predictors are entered into the model first to control for possible bias due to design characteristics. The variables class size, instructional time, and age are entered into the equation secondly to remove significant variation between outcomes due to these educational setting variables. Finally, the predictors of the instructional subset are entered. To retain power, one predictor of each subset is entered into the equation each time. Significant predictors of each subset are included in the regression model to analyze whether the predictors of subsequent subsets explain additional variance. The significance level for each predictor is set at $\alpha = .05$. All analyses are performed with VKHLM (Bryk, Raudenbusch & Congdon, 1994).

## Results

None of the methodological predictors reduced heterogeneity significantly. This implies that there is no systematic effect of any methodological characteristic on treatment outcome. Of the educational setting variables, class size produced a small, but significant negative effect on the treatment outcome, i.e. instruction in smaller group settings results in larger effect sizes. After inclusion of class size, only the dummy variable 'clue instruction' from the instructional subset explained additional variance. Clue instruction shows a significant positive relation with treatment outcome (see Table 4).

The variable 'class size' reduces the variance by 24.4%, and after inclusion of both class size and clue instruction in the regression equation, 58.8% of the variance is explained. The homogeneity test hereafter no longer indicates heterogeneity ($p = .17$).

### *Sensitivity Analysis*

Sensitivity analysis is performed to evaluate possible threats to the conclusions of the meta-analysis. Applying the leave-one-out method does not seriously affect the results of the random effects model. Leaving out one treatment each time, causes the generalized effect size $\Delta$ to range from a low .39 to a high .46, which does not deviate markedly from the outcome of .43 that is computed over all treatments. The upper and lower boundaries of the 95%-confidence intervals fall within a slightly wider range from .22 to .65 now (the 95%-confidence interval was .25-.62 for all treatments). The outcome of the meta-analysis does not seem to be heavily influenced by particular studies.

Another threat to the conclusions of meta-analysis is publication bias, especially in a field where many small-scale studies are being published (Begg, 1994). A serious effect of publication bias is unlikely in this case, however, as the proportion of not-published studies is relatively large in this meta-analysis. Other analyses do not give an indication of publication bias either. A plot of sample size versus effect size shows a funnel-like shape, which is indicative of no bias (Light & Pillemer, 1984). Furthermore, Orwin's fail-safe *N* (Orwin, 1983) is 24.58 for a small effect size, which means that 25 (not-published) studies should report zero-outcomes to reduce the overall effect size to a nonsignificant outcome.

A strong effect of bias due to selection factors seems unlikely also, although included treatments show more significant results than not-included studies. Eleven of the 21 included treatments showed significant results (52%), whereas for the not-included treatments with a control group, four out of 11 treatments reported significant outcomes (36%), while the group sizes (respectively 62.3 and 61.3) and class size are comparable (respectively 17.6 and 18).[3]

## Discussion

The field of vocabulary instruction is characterized by different approaches to the extension of the reading vocabulary of students. The four main instructional types that were distinguished by Beck and McKeown (1991) contribute in different ways to vocabulary growth. As research has shown, direct instruction of word meanings is beneficial for poor and average readers (Stahl & Fairbanks, 1986; Marmolejo, 1990). Wide reading also has its value, as research of the incidental word learning hypothesis has shown that readers gain knowledge of a small, but unnegligible proportion of the unknown words they encounter (see Nagy, Anderson, & Herman, 1987). Learning how to consult a dictionary is another essential skill for readers. Finally, it makes sense to teach students how to derive word meaning from context. As this meta-analysis shows, deliberately deriving word meaning from context is amenable to instruction and the effect of even relatively short instruction is rewarding. The mean size of the instructional effect is 0.43 standard deviation units, which is close to an effect of 'medium' size, as defined by Cohen (1988). In evaluating the outcome of this meta-analysis, it must be taken into account that in many studies experimenter-developed tests were used, and effects may be smaller with standardized tests. The positive outcome of this meta-analysis, however, is unexpected, considering the predominantly cautious remarks in the literature about the effects of instruction in deriving word meaning from context.

Our estimate of the effects that are achieved with instruction in the skill of deriving word meaning from context can be interpreted by comparing it to the rate of natural growth with age as investigated by De Glopper, van Daalen-Kapteijns, and Schouten-van Parreren (1997). In a cross-sectional study, the skill in deriving word meaning of students in Grade 6, 8, and 10 were investigated. A difference of 0.46 standard deviation units was found between the achievement scores of students from Grade 6 and 8, and a difference of 0.74 standard deviation units between Grade 6 and 10. The population of this study is roughly comparable to the population of this meta-analysis. The mean effect size of 0.43 after instruction can therefore be cautiously interpreted as the difference that would be found after a period of two years of natural development.

### *Instructional Implications*

It is complicated to formulate directions for educational practice of deriving word meaning from context. The empirical evidence is not unequivocal and the theoretical foundations of instruction are sparse or even absent. Different researchers have tried to establish some empirical base of context clue instruction by investigating the frequency of clue types in natural text and their general 'success rates', but this line of investigation has never been worked out fully. We do not know of any theory-embedded research of cloze instruction to enhance the ability to derive word meaning from context. The research of strategy instruction lacks a process model that describes the process of deriving word meaning from context of good and poor readers. Such a cognitive process model would be of help, in conjunction with other tools, in the design of strategy instruction. Although the process of deriving word meaning from context has been investigated in some think-aloud studies (Van Daalen-Kapteijns & Elshout-Mohr, 1981; Van Daalen-Kapteijns, Elshout-Mohr, De Glopper, & Schouten-van Parreren, 1997; Werner & Kaplan, 1952), we do not know yet if any expert strategy exists that can be taught to novices with some success. Recent studies, however, make reference to process models to motivate choices in the instructional design (see Van Daalen-Kaptejins et al., 1997b; Goerss, 1995). In short, the research of instruction in deriving word meaning from context is still in its infancy, although recent studies are beginning to explore more theory-embedded experimentation by combining the growing insights in the field of vocabulary skills.

The present investigation provides some suggestions for effective instruction. The differences in study outcomes are large, and not every instruction appears equally successful. Cloze instruction does not seem a very interesting method to explore in future research at this point. Although practicing cloze tests can certainly be of use, an exclusive reliance on this type of exercise seems of limited value for a number of reasons. First, the studies investigating the effects of cloze instruction do not show major improvements with the exception of Sampson, Valmont, & Allen (1982). Second, the students in the cloze studies have mainly practiced cloze tests and have received relatively little explicit instruction in how to use context. Future studies should therefore investigate whether cloze practice can be supplemented with direct instruction. However, this line of research is seriously hindered by a limitation of the cloze test. Students that practice this test can fill in only words they already know. It is a drawback of cloze instruction that it is not clear how practicing cloze tests can be used to enhance the ability of deriving the meaning of new concepts.

The result of the multilevel regression analysis suggests that clue instruction is superior to other instruction types. This analysis is exploratory and, hence, its outcomes need to be interpreted with some caution. Although the ratio of significant predictors in the model and experimental treatments is 2 : 21, these predictors are selected from a larger set, and the results may have been influenced by capitalization on chance (Bryk & Raudenbusch, 1992). Furthermore, it is not clear whether clue instruction leads to a specific improvement in tasks that include the specific clues instructed only, or that a broader effect may be expected, because it could not be deduced from the reports whether the posttests included items that contained the instructed clue types only. The possibility is therefore not excluded that the selected clues are amenable to instruction but that no transfer occurs to other clue types or to contexts that do not have specific clues. If no transfer occurs, then the effect of clue instruction is very specific and less encouraging. The fact is that the authors of even the earliest clue classifications have noticed that the distinct clues appear not very often in natural text. It is, however, an interesting question whether starting with explicit clues and helpful contexts is a very effective method to help younger readers to discover the aid of context and to develop gradually a general ability in deriving word meanings.

A related issue that should be addressed in future research is how many and which clues should be instructed. Instruction in experimental research has so far only been conducted on a selection of clue types that stem from different clue classifications. None of these classifications is based on a leading principle that divides up the domain of clues exhaustively into mutually exclusive types. Because the taxonomic quality of the clue classifications is poor, students may experience difficulties if instruction is extended to a broader range of clue types.

Strategy instruction of word derivation abilities too seems a promising method to explore in future research. The positive results that are achieved with some instructions are interesting because students are instructed in a general strategy that is applicable to a wide range of contexts. Not all treatments, however, are successful. The outcomes of the study of De Glopper et al. (1997) and the 'low practice' treatment of Jenkins et al. (1989) are disappointing. The 'enriched context' treatment of Kranzer (1988) even produced negative outcomes on both word meaning derivation from context and incidental word learning measures. Also the difference in outcomes between the 'rule plus practice' treatment of Carnine et al. (1984) and the 'practice-only' treatment is negligible small ($d = .08$) and, hence, no superior effect of strategy instruction was demonstrated. The absence of a significant difference favoring the 'rule plus practice' group may lie in the superficiality of the rule that was provided to them ("When there's a hard word in a sentence, look for other words in the story that tell you more about that word", p. 197). Friedland (1992) did not find significant differences either between the strategy instruction and the clue instruction group. The differences on an immediate and delayed test, favoring the strategy treatment, were small and nonsignificant (*d* = .12 and .33 respectively).

Kuhn and Stahl (1998) concluded that instruction and practice seem equally effective, based on their finding that all four studies that included a practice-only condition (Carnine et al., 1984; Patberg et al., 1984; Schwartz & Raphael, 1985, Study 2; Sternberg, 1987, Study 2) reported no significant differences between this condition and an instructional condition. In addition, they stated that '(i)n all four studies, the practice-only treatment significantly outperformed a control group' (p. 129). However, their vote-counting analysis of these studies and the following conclusions raise some issues.

First, the statement that in all four studies the practice-only treatment significantly outperforms the control group does not seem correct. Only Carnine et al. report a significant difference between the practice-only condition and the control group, whereas in the other studies no statistically significant difference can be demonstrated between these two conditions. Performing vote-counting for the four studies leads therefore to a 4 : 0 score favoring instruction, and performing vote-counting on the practice-only conditions shows a 1 : 0 score, comparing them with the control group. Instruction produces a greater and more reliable effect than just practice. Seen from this perspective, the tentative conclusion of Kuhn and Stahl that instruction and practice-only seem equally effective does not seem warranted.

Kuhn and Stahl (1998) compared the instructional treatment and the practice-only condition directly in their vote-counting analysis. Their conclusion that no significant difference is found between instructional treatments and the practice-only condition is not correct for the study of Sternberg (1987), since he reports that 'the training groups showed significant greater gains than did the control groups' (p. 103) that included a practice-only condition. Finally, vote-counting is not the best way to analyze results from experimental studies because nonsignificant differences go undetected. Unfortunately, effect sizes can be computed for Carnine et al. (1984) and Schwartz and Raphael (1985) only. Comparing the effect sizes for the instructional treatments and the practice-only conditions with each other, the difference found in Carnine et al. seems negligible, whereas the difference in Schwartz and Raphael does not (see Table 2).

### *Methodological Implications*

A few suggestions for future research on the instruction of word derivation can be formulated. First, accumulation of evidence in the field of word meaning derivation would be served by a higher methodological quality of the studies and their reports. Older as well as more recent studies could not be included in the meta-analysis because the basic statistics necessary to compute effect sizes were not provided.

Furthermore, it is important from a methodological point of view, that in future studies an untreated control group design with pretest and posttest is applied with random assignment of students to conditions. Internal validity could be increased further by taking statistical power into consideration. The sample size of the word derivation studies conducted is small. Assuming an average effect size of .43, the statistical power of the 'average study' included in the meta-analysis is only .52, using a one-sided t-test at the significance level of .05. This means that only half of the studies will produce statistically significant results if an effect of this size is present, a number that corresponds to the 11/21 ratio of significant results. A total sample size of 136 would be needed for a statistical power of .80 (statistical power can also be raised by using covariance analysis). Seen from this perspective, the differences between treatments of some multiple-treatment studies seem intuitively too subtle to expect statistically significant differences. It is further interesting to include process measures of actual strategy use more often (see also Lysynchuk, Pressley, d'Ailly, Smith & Cake, 1989) to investigate how instruction affects students strategies.

It is still an open issue what the effects of instruction in deriving word meaning are for incidental word learning. Patberg et al. (1984) and Kranzer (1988) stand out as the two experimental studies that have addressed this hypothesis so far. Instruction yielded a small nonsignificant effect on incidental word learning, and more research is needed to establish a firm empirical base for this hypothesis. The correlation of *r* = .66 between deriving abilities and incidental word learning found by Kranzer, suggests that these two abilities are related only to some extent.

Effects of word meaning derivation instruction on incidental word learning may be smaller. Deriving word meaning from context is a deliberate and intensive process and can therefore be considered as the maximum performance (see also the distinction between 'maximum' and 'typical performance' of Shefelbine, 1990). During reading under normal conditions, a reader may not be oriented at deriving the meaning of unknown words (see Barnes, 1986; Stallman, 1991; Stanley, 1989), and if incidental word learning occurs during normal reading, probably less time and effort will be spent in it. In addition, incidental word learning is a more complex task that involves besides meaning derivation also the memorization of the form and derived meaning of the word. It seems therefore plausible that instruction should be delivered for a relatively long period before a significant transfer effect from intentional word learning to incidental word learning can be expected.

Results of instruction would be even more convincing if improvement of both deliberately deriving word meaning from context and incidental word learning abilities is demonstrated (see also Kuhn & Stahl, 1998). Future studies should therefore incorporate a measure to evaluate the transfer of deriving word meaning to incidental word learning abilities. Future experimental studies must bridge the gap between word meaning derivation and incidental word learning to investigate their assumed relationship that creates the "sound and persuasive rationale" (Jenkins et al., 1989, p. 218) that underlies or at least partly motivates word derivation studies.

## Notes

[1] The standard deviations of the posttest scores of Guarino (1960; the Michigan study) were computed from the raw scores in the appendix.

[2] Studies that combined clue instruction with strategy instruction were coded as clue instruction. Although the classification of Buikema and Graves (1993) is accurate given this classification system used, it does not take into account that considerable

emphasis is given to teaching the students a strategy. In this study, the teacher read students a definition about clues which pertain to the senses, actions or functions of a word, which was briefly discussed. Later, however, instruction in this study focused on a four-step strategy that received considerably more emphasis (1. Box in the unknown word and write the word below the passage, 2. List words and phrases which are cues to the possible meaning, 3. Think about what the word might mean, considering past experience, the part of speech of the word, what the word can not be, and what it might be, and 4. Guess what the unknown word means.).

[3] The not-included studies with a control group are Butler (1943), Peterson (1942), Jensen (1943) (all three of them cited in Guarino, 1960); Askov & Kamm (1976), Farr (1987), Hafner (1965), the 'active teaching' treatment and the 'practice only' treatments of Patberg et al. (1984), Patberg & Stibbe (1985; cited in Beck & McKeown, 1991), 'Experiment I' of Schwartz & Raphael (1985), and 'Experiment I' of Sternberg (1987).

## References

*(References marked with an asterisk indicate studies included in the meta-analysis.)*

Artley, A.S. (1943) Teaching word meanings through context. *Elementary English Review, 20,* 68-74.

Askov, E. N., & Kamm, K. (1976). Context clues: Should we teach children to use a classification system in reading? *Journal of Educational Research*, 69, 341-344.

Barnes, J. A. (1986). Reading comprehension: the role of schemas and purpose in learning from context. *Dissertation Abstracts International, 47(04).* (University Microfilms No. AAC86-14719).

Beck, I. L., & McKeown, M. G. (1991). Conditions of vocabulary acquisition. In R. Barr, M. L. Kamil, P. B. Mosenthal & P. D. Pearson (Eds.), *Handbook of Reading Research* vol. II (pp. 789-814). New York: Longman.

Beck, I. L., McKeown, M. G., & McCaslin, E. (1983). All contexts are not created equal. *Elementary School Journal, 83,* 177-181.

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *Handbook of Research Synthesis* (pp. 399-409). New York: Russell Sage Foundation.

* Bissell, L. Vandermer (1982). *Training with forced-choice cloze tasks.* [Place]: University of Michigan.

Boettcher, J. V. (1980). Fluent readers' strategies for assigning meaning to unfamiliar words in context. *Dissertation Abstracts International 41*(03). (University Microfilms No. AAC80-19516).

Butler, H. (1943). *Finding word meanings from context in grades five and six.* Boston University, unpublished master's thesis.

Bryk, A. S., & Raudenbusch, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park: Sage Publications.

Bryk, A. S., Raudenbusch, S. W., & Congdon, R. T. (1994). *HLM 2/3. Hierarchical linear modeling with the HLM/2L and HLM/3L programs.* Chicago: Scientific Software International.

* Buikema, J. L., & Graves, M. F. (1993). Teaching students to use context cues to infer word meaning. *Journal of Reading, 36(6),* p. 450-457.

* Carnine, D., Kameenui, E. J., & Coyle, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly, 19(2),* 188-204.

Carr, E. M., & Mazur-Stewart, M. (1988). The effects of the Vocabulary Overview Guide on vocabulary comprehension and retention. *Journal of Reading Behavior, 20(1),* 43-62.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty in the text: Implications for instruction. *Journal of Reading Behavior, 26(4),* 413-437.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum.

Conger, A. J. (1980). Integration and generalization of Kappas for multiple raters. *Psychological Bulletin, 88(2),* 322-328.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation; Design and analysis issues for field setting.* Boston: Houghton Mifflin.

Cooper, H., & Hedges, L. V. (1994). Potentials and limitations of research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 521-529). New York: Russell Sage Foundation.

* Cox, J. A. K. (1974). A comparison of two instructional methods utilizing the cloze procedure and a more traditional method for improving reading. *Dissertation Abstracts International 35(10).* (University Microfilms No. AAC75-9580).

Daalen-Kapteijns, M. van, & Elshout-Mohr, M. (1981). The acquisition of word meanings as a cognitive learning process. *Journal of Verbal Learning and Verbal Behavior, 20,* 386-399.

Daalen-Kapteijns, M. van, Elshout-Mohr, M., De Glopper, K., & Schouten-van Parreren, C. (1997). *Natuurlijke strategieën bij het afleiden van woordbetekenissen uit context [Natural strategies of deriving word meaning from context].* (rep. no. 464). Amsterdam: SCO-Kohnstamm Institute.

Daalen-Kapteijns, M. van, Schouten-van Parreren, C., & De Glopper, K. (1993). Het afleiden van woordbetekenissen uit de context [The derivation of word meanings from context]. *Levende Talen, 485,* 589-593.

* Daalen-Kapteijns, M. van, Schouten-van Parreren, C., & De Glopper, K. (1997). *The training of a word learning strategy: results in process and product* (Rep. No. 463). Amsterdam: SCO-Kohmstamm Institute.

D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior, 23(1),* 109-122.

Farr, C. W. (1987). *Effects of inferencing training on verbal abilities and mathematics problem-solving among adult basic education students.* [Place]: University of Wyoming [dissertation, UMI].

Friedland, E. S. (1992). *The effect of context instruction on the vocabulary acquisition of college students.* Buffalo: State University of New York.

Frye, S. (1975). Training retarded and normal pupils to use context clues to derive word meanings. *Journal of Research and Development, 8, Monograph,* 66-68.

Gifford, A. P. (1993). *An investigation of the effects of direct instruction in contextual clues on developmental reading students' ability to increase vocabulary and reading comprehension scores.* Carbondale: Southern Illinois University [dissertation, UMI].

Glopper, K. de, Daalen-Kapteijns, M. van, & Schouten-van Parreren, C. (1997). *Vocabulary knowledge and skill in inferring word meaning from context* (Rep. No. 462). Amsterdam: SCO-Kohnstamm Institute.

Goerss, B. L. (1995). *Study to train elementary students to become more sensitive to context clues.* ERIC Document Service ED 392033.

Graves, M. F. (1986). Vocabulary learning and instruction. In E. Z. Rothkopf & L. C. Ehri (Eds.), *Review of Research in Education*, vol. 13, pp. 49-89. Washington, DC: American Educational Research Association.

* Guarino, E. A. (1960). *An investigation of the effectiveness of instruction designed to improve the reader's skill in using context clues to derive word meaning.* [Place]: Syracuse University [dissertation, UMI].

Hafner, L. E. (1965). A one-month experiment in teaching context aids in fifth grade. *Journal of Educational Research, 58(10),* 472-474.