Review of Educational Research
Fall 1999, Vol. 69, No. 3, pp. 261-285

Incidental Word Learning While Reading: A Meta-Analysis

M.S.L. Swanborn
K. de Glopper
SCO-Kohnstamm Institute,
University of Amsterdam

A meta-analysis of 20 experiments examining incidental word learning during normal reading shows that students learn around 15% of the unknown words they encounter. A test of homogeneity indicates that study outcomes diverge, although their heterogeneity is relatively small (19%). An exploratory multi-level analysis of the variability in the results suggests that several factors affect the probability of learning an unknown word while reading: pretest sensitization, students' grade level, students' level of reading ability, the sensitivity of assessment methods to partial word knowledge, and the amount of text surrounding the target words. A model that contains students' grade level and assessment methods' sensitivity to partial word knowledge predicts 66% of the systematic variance in the effect sizes. Implications for research and instruction are discussed.

For decades, it has been a common assumption among educational researchers that words are learned incidentally during reading (Elivian, 1938; Hafner, 1932). Written and oral contexts are supposed to be the major sources of vocabulary growth (Nagy & Herman, 1987; Sternberg, 1987). However, the claim that many if not most words are learned from context was unsubstantiated until about a decade ago, when Nagy, Herman, and Anderson (1985) examined it and concluded that children indeed do derive and learn new vocabulary incidentally while reading. The probability of success in incidental learning of meanings of unknown words under natural reading conditions was estimated to be five percent by Nagy, Anderson, and Herman (1987).

Although the probability of five percent has been cited by many researchers (e.g., Kranzer, 1988; Shefelbine, 1990) this chance of learning a new word incidentally has been the subject of discussion. The results of other studies on incidental word learning vary from about 5 percent to 15 percent (Herman, 1985). It is an interesting to question what the average probability is and why there is such a difference between probability estimates.

Incidental word learning studies differ on several characteristics such as texts, words, and designs used. The variation in word learning chance found across

studies can perhaps be attributed to one or more of these study features. Granick (1997), in a narrative review, resumes the studies on incidental word learning and recognizes the different probability figures. He does not, however, provide us with an estimate of the average word learning probability, nor with a systematic explanation for the different figures. Such an explanation might reside in the fact that past studies used different designs, different subjects, different tests, and different texts. It may well be that these differences are related to the variation in incidental word learning chance.

In order to estimate the average probability and to analyze the extent to which differences between study outcomes are due to systematic variation or to sampling error, we performed a statistical meta-analysis (Glass, McGaw, & Smith, 1981). This kind of analysis may clarify the differences in probability figures between studies on incidental word learning. A notable advantage of meta-analyses over narrative reviews is the possibility of exploring variables that may not be of statistical significance in individual studies—and are therefore often viewed as not relevant—but may well be of significant relevance when the analysis is based on the combined samples of separate studies.

In order to be able to perform this meta-analysis, it is necessary to formulate exactly what incidental word learning means. We define incidental word learning as *the incidental, as opposed to intentional, derivation and learning of new word meanings by subjects reading under reading circumstances that are familiar to them.* The word *incidental* implies that the purpose for reading does not specifically provoke learning or directing attention to the meaning of unknown words. Before reading no mention has been made of any assignment or intent of learning new words. In the literature, researchers also used the term *casual word learning* (Hafner, 1932). We prefer the term *incidental*, because it reflects more obviously the opposite to intentional learning and we will use it throughout this article.

In the definition, the terms *derivation* and *learning* have been included on purpose, since earlier studies often claim to assess incidental word learning, while they are in fact assessing only the intentional deriving ability (Carnine, Kameenui, & Coyle, 1984). If one asks a student to figure out the meaning of unknown words while reading a text, one calls upon the skill to derive word meanings in an intentional way. The derivation of word meanings is involved in incidental word learning, but it is not triggered by the purpose of the reading task. Next to the process of deriving word meanings, incidental word learning also involves the retention of word meanings in memory. What is remembered of the unknown words in a text after reading? Are any of the derived word meanings remembered? It takes both successful derivation and memorization to expand one's vocabulary.

Familiar reading circumstances is included in the definition to ensure that the research estimates the real-life chance of incidentally learning word meanings, which implies the use of texts subjects usually encounter in or outside the classroom. Thereby they will not specifically focus their attention on the vocabulary. Provided that incidental word learning is measured under these circumstances, one can assess the effectiveness of learning words while reading.

The definition corresponds to some conditions incidental word learning studies have to fulfil, as formulated by Nagy et al. (1985). In order to assess as

validly as possible the contribution of incidental vocabulary learning to vocabulary growth, studies should be performed under as natural a reading circumstance as possible; that is, studies have to simulate normal reading conditions. According to Nagy et al., subjects must not be aware of the focus of the study; they have to read while no reading purpose is stated or a reading purpose is stated that is familiar to them. Secondly, subject must read authentic texts that are not specifically written for the study, and thirdly, next to unknown words representing familiar concepts, knowledge of completely new concepts also has to be assessed. In his review, Granick (1997) mentions eight studies measuring incidental vocabulary acquisition according to the conditions formulated by Nagy et al. (1985). Studies like Jenkins, Stein, and Wysocki (1984) have been left out of his overview, because the authors used specifically constructed or modified texts. However, we do not think that all studies using modified texts should necessarily be left out. First of all, students often read schoolbook texts that are especially written to serve learning purposes. This means that reading new, specifically constructed texts is a natural reading activity for them. Secondly, we do not think that modified texts necessarily make context more transparent or relationships in the text more obvious. Whether context is made more considerate or not will depend on the specific changes made to the text. One may imagine that some contexts are made less, and others more transparent, such as in Diakidoy's study (1993).

With regard to Nagy et al.'s (1985) third condition we note that there are natural texts which do comprise new concepts, as well as texts that do not. A student will encounter both during his reading activities; we therefore accept studies of incidental word learning that use texts with or without new concepts.

In a general sense three groups of incidental word learning studies may be distinguished. The first group of studies (Granick, 1997; Nagy et al., 1985, 1987; Shu, Anderson, & Zhang, 1995; Stallman, 1991) sets out to measure the amount of incidental word learning. The authors all use a design in which two groups are given two different texts to read, but are tested on knowledge of unknown words from both texts. The influence of different factors such as level of reading ability, age group, and conceptual difficulty of the words is examined. The results of the studies in the first group vary. Sometimes an effect of reading ability is demonstrated (Herman, 1985; Konopak, 1988b); sometimes no significant relationship between reading ability and incidental word learning can be detected (Nagy et al., 1987). The same goes for the other variables in these studies. Although all these studies use the same design and sometimes even the same texts, differences exist in the way incidental word knowledge is measured. Multiple choice tests are used, but also definition tasks or interviews. Sometimes credit is given for partial word knowledge, but sometimes full word knowledge is expected. Tests are mostly administered immediately after reading, but in one study the test was administered seven days afterwards.

The second group of studies (Diakidoy, 1993; Gordon, Schumm, Coffland, & Doucette, 1992; Herman, 1985; Konopak, 1988a, 1988b; Konopak et al., 1987) sets out to examine the effect of textual manipulations on incidental word learning. These studies all use a pretest-posttest design, except for Herman who uses the design adopted by the studies from the first group. They all demonstrate significant effects of considerate versus inconsiderate contexts on incidental

word learning. They differ, however, in reading circumstances. Among other things, different reading purposes are given to the students.

The third group comprises a variety of studies with different rationales. Stahl (1989) is interested in the effect of pre-teaching topic-related information on incidental word learning, but such effects are not found. Kranzer (1988) sets out to examine the influence of an instruction in deriving ability on incidental word learning. She gives a measure of incidental word learning for the control group merely as a by-product. Durkin (1990) wants to see what multiple encounters with a word mean for incidental word learning. Schwanenflugel, Stahl, and McFalls (1997) compare the amount of learning relative to previously unknown and partially known words. This last group of studies differs from the other groups, and also among the studies themselves, on design factors, subject factors, assessment factors, and material-related factors.

From the above it becomes clear that the results and the characteristics of incidental word learning studies vary considerably. It would be interesting to estimate the average word learning probability and to examine the relation between these factors and the probability of learning new word meanings incidentally. In the remainder of this article we review the studies on incidental vocabulary learning while reading by means of a meta-analysis. We will try to relate study outcomes to factors supposedly influencing the probability of learning words incidentally while reading.

Method

Inclusion Criteria

Incidental word learning is defined as the incidental, as opposed to intentional, derivation and learning of new word meanings by subjects reading under reading circumstances with which they are familiar. This definition implies a number of criteria a study must meet in order to be included in the meta-analysis:

- Incidental word learning studies must assess incidental word learning from reading in the mother tongue. Studies dealing with reading in a second language are excluded.
- No attention must be drawn to the vocabulary in the text: Studies using underlined words are excluded.
- In the purpose stated to the subjects before reading a text, no mention must be made of presenting or learning new vocabulary.
- Texts may be newly constructed, provided that context is not intentionally made more transparent than it would be in natural text. The same goes for existing texts.
- Studies can only be included if the subjects encounter the word in only one text. The words may appear more than once in the same text, for this cannot be controlled, but if subjects encounter the same word in more than one text in a relatively short time span, their attention might be artificially drawn to the word, and it will not be possible to assess the degree of knowledge students incidentally gain after encountering the word in one text.

An additional criterion for inclusion of studies pertains to the availability of statistical information. Studies have to provide enough information to allow calculation of the probability of learning an unknown word.

Sample of Studies

We consulted multiple databases in order to find studies assessing incidental word learning: ERIC (1965- December 1997), PsycLit (1974- September 1997), Linguistics and Language Behavior Abstracts (1973- July 1997), Dissertation Abstracts (1988- August 1997), and Current Contents on Disc (until September 1997). With regard to the databases ERIC, LLBA, and PsycLit, keywords were chosen only after entering the study of Nagy et al. (1985). This resulted in the following useful keywords concerning ERIC: vocabulary development, word recognition, and word study skills. PsycLit and LLBA were searched using combinations of '[incidental learning, context, and reading]'. Dissertation Abstracts was searched using combinations of '[incidental learning, education and reading, context, and vocabulary acquisition]'. These keywords were identified by means of the accompanying thesaurus. Finally, we searched CCOD with the terms '[incidental, context, and reading]'. We completed the search for relevant studies by following up on references and suggestions of colleague researchers and reviewers. The studies of Schwanenflugel et al. (1997), Stahl (1989), and Stallman (1991) were found this way.

The computer search produced a large number of studies examining incidental word learning. Not all identified studies met our criteria of inclusion. A number of studies involved students of English/French as a second language (Dupuy & Krashen, 1993; Fischer, 1994; Huckin & Zhendong, 1986; Mondria & Wit-De Boer 1989). Several other studies involved listening instead of reading (Eller, Pappas, & Brown, 1988; Elley, 1989; Sénéchal & Cornell, 1993). These studies are beyond the scope of meta-analysis. In case their numbers are sufficient, a separate meta-analysis of studies in these fields might be of interest.

In other studies, students' attention was drawn to the target words by underlining these words (Carnine et al., 1984; Hafner, 1932; Schatz & Baldwin, 1986) or by familiarizing the students with the words in advance (Jenkins et al., 1984). On closer inspection, other studies proved to assess intentional instead of incidental learning (Browne, 1989; Elivian, 1938; Shefelbine, 1990). One study lacked a control condition necessary to estimate an effect size (Bonacci, 1993). This study neglected to control for any pre-existing knowledge of the target words. The studies of Stanley and Ginther (1991) and Jenkins et al. (1984) calculated the word learning probability only after multiple encounters with the target words in multiple texts (e.g., for each target word the student reads more than one text).

Three studies could not be obtained: Caracciolo, Fabio, and Trombetta (1988), Martin-Rehrmann (1990), and Stein (1989). They were not available from Dissertation Abstracts International.

All in all, 15 studies are suited and available for inclusion in the meta-analysis. The studies of Diakidoy (1993) and Shu et al. (1995) report about two different experiments examining the incidental word learning probability. In the studies of Gordon et al. (1992) and Konopak (1988a, 1988b) the same experiment is performed among two different reading ability groups. In total, 20

experiments are available. They provide us with enough substantial information to be considered an appropriate assessment of the incidental word learning probability.

One could be concerned about the relatively small number of data points (20) in the meta-analysis. Rosenthal (1995), however, states that meta-analytic procedures can be applied to as few as two studies. The only danger involved with small sample sizes is the relative instability of the results. With respect to the stability of the estimation of the mean effect size, according to Hedges (1994), within-study sample sizes have to be taken into account, too. The retrieved sample of studies as a whole represents 2130 subjects, with an average of 107 subjects per experiment. When it comes to explaining between-study heterogeneity in effect sizes, small numbers of studies limit the power of regression analyses (Cohen, 1988). Care should therefore be taken not to adopt a too conservative method of analysis.

Computation of Effect Size

The probability of learning an unknown word incidentally while reading is defined by Nagy et al. (1985) as the increase in the number of words known to a given criterion divided by the number of words originally not known to that criterion. Only the studies of Granick (1997), Nagy et al. (1985), Nagy et al. (1987), Shu et al. (1995), and Stallman (1991) that were specifically designed to assess the amount of incidental word learning and the study of Herman (1985) report estimates of this probability; the other studies do not. Probabilities could however be computed for all other studies by using the available statistical information. In cases where a pretest was used to assess prior word knowledge we considered the results on the pretest as those from the control condition; we subtracted the mean on the pretest from the mean on the posttest and divided the result by the maximum score minus the mean on the pretest. For two studies (Kranzer, 1988; Stahl, 1989), this was not possible because of the absence of a pretest. However, in the Kranzer study a pretest assessing target word knowledge was administered to a comparable group of students. These results were available and used as the control condition. In the Stahl study, two groups read the same normal passage, but in only one passage the target words were present. There were synonyms of easier words in the other text that served as a control. The scores of the group that read the control text were subtracted from the scores of the experimental group and divided by the maximum score minus the mean of the control group.

The Durkin (1990) study tested groups separately on knowledge of simple synonym words and concept challenge words. Since the groups are comparable and the experiment is the same, we collapsed these groups into one group. The same was done for the study of Stallman (1991), where the conditions 'Read the text' and 'Read the text. Try to understand the important ideas' were taken together. The studies using multiple grades (Nagy et al., 1987; Shu et al., 1995) were also averaged into one group, since separate probability estimates for each grade could not be obtained.

The studies by Durkin (1990), Granick (1997), and Nagy et al. (1985) used two different measures. Since one measure always followed the other in time, and an effect of the first assessment on the second assessment was not to be

TABLE 1
Number of subjects, probability of learning an unknown word (Prob), effect size (Logit (p)), and effect size variance of the studies included in the meta-analysis

Study	*N*	Prob	Logit (*p*)	Variance
Diakidoy, 1993, exp.1	50	.10	-2.22	.23
Diakidoy, 1993, exp.2	73	.18	-1.50	.09
Durkin, 1990	54	.06	-2.70	.31
Gordon et al., 1992, H	19	.14	-1.82	.44
Gordon et al., 1992, A	19	.03	-3.55	1.93
Granick, 1997	349	.06	-2.77	.05
Herman, 1985	39	.13	-1.67	.19
Konopak, 1988a, H	27	.54	.16	.15
Konopak, 1988a, A	28	.42	- .29	.15
Konopak, 1988b, H	27	.35	- .61	.16
Konopak, 1988b, A	25	.17	-1.55	.28
Konopak et al., 1987	21	.27	-1.01	.24
Kranzer, 1988	19	.21	-1.35	.32
Nagy et al., 1985	57	.11	-2.13	.19
Nagy et al., 1987	352	.05	-2.98	.06
Schwanenflugel, 1997	33	.12	-1.97	.28
Shu et al., 1995, exp.1	146	.10	-2.23	.08
Shu et al., 1995, exp.2	301	.09	-2.38	.04
Stahl, 1989	391	.13	-1.92	.02
Stallman 1991	100	.23	-1.23	.08

excluded—for example, when a definition task follows a multiple-choice test—we calculated a probability estimate for the first assessment only. The probabilities of learning unknown words while reading that we obtained for each of the 20 experiments are reported in Table 1.

Bryk & Raudenbusch (1992) and Rosenthal (1994) provide us with different effect size indicators as used in meta-analyses, such as *r* (population correlation between variables *X* and *Y*), Hedges's *g* (difference between population means divided by average population standard deviation), and logit *d'* (difference between logit transformed population proportions). In our meta-analysis the probabilities were converted to effect sizes by treating them as proportions and by subjecting the proportions to a normalizing logit transformation. This transformation will often improve the tenability of the normality and variance-known assumptions (Bryk & Raudenbusch, 1992: 169), especially in meta-analytic applications. For our data, such a transformation is necessary: the values of proportions are bound to be between 0.00 and 1.00 and near their extremes of 0.00 and 1.00 changes in proportions are hard to produce. The following logit function that transforms the observed proportions to values from -infinite to +infinite was therefore used:

$$f(p) = \ln(p/(1-p))$$

(Bryk & Raudenbusch, 1992; Hox, 1995; Hox & de Leeuw, 1997). The variance

of the resulting effect sizes is known to be:

$$1/(Np(1\text{-}p))$$

(Hox & de Leeuw, 1997). The original probabilities and the transformed effect sizes are highly correlated (r = .95).

Coding

To account for possible differences between study outcomes, studies were coded according to four groups of factors: Study conditions, subject factors, assessment factors, and material-related factors. The group of study conditions was composed of experimental design, pretest sensitization, time interval between pretest and reading, and time interval between reading and the posttest.

Experimental design was chosen to be coded in order to identify the studies using gain scores as opposed to studies administering no pretest to the subjects in the study (Granick, 1997; Kranzer, 1988; Stahl, 1989). We expected studies using a pretest or proxy to yield a more accurate assessment of the word learning chance than studies that failed to administer one. Experimental design was coded as untreated control group design without pretest versus untreated control group design with pretest or proxy.

Although not all studies used a pretest to measure existing knowledge of the target words, the studies that did sometimes tested the target words only. Others administered long forms composed of the target words and a series of distractor items. We expected the pretest with target words only to sensitize the subjects to the words and provoke recognition later on when reading the text. The effect size would consequently be overestimated. Pretest sensitization was coded as absence of distractor items versus presence of distractor items.

The time interval between pretest and reading of the text was coded in order to find out whether a pretest testing knowledge of the target words just before reading the text would draw too much attention to the target words in the text. Thus, the effect size would be overestimated. The time interval was coded by taking the number of days between the pretest and the reading.

The time interval between reading of the text and the posttest of the knowledge of the target words was coded similarly. It might very well be that subjects recall some of the words immediately after reading the text, but that they have forgotten many of the words, if not all, a week later. There might therefore be differences between studies due to different moments of assessing target word knowledge. We expected that studies testing immediately after reading would yield larger effect sizes.

Subject factors comprise grade level and level of reading comprehension. Mayberry, Taylor, and O'Brien-Malone (1995) showed that older children outperform younger ones on explicit as well as implicit learning tasks. Incidental word learning can be seen as an implicit learning task. Differences in study outcomes related to grade could point to the absence or presence of reading strategies. We expected that students in the early grades learn less words incidentally than those in the higher grades, because older students will have been exposed to and learned more reading strategies at school. This was also an open-ended coding variable. The different grades present in the studies of Nagy

et al. (1987) and Shu et al. (1995) were averaged.

Next, we chose to code the level of reading comprehension. This factor might well relate to differences between poor and good readers. Elivian (1938) found a large difference in the capacity to derive unknown word meanings between high- and low-ability readers, which is part of learning word meanings incidentally. Among other researchers (e.g., McKeown, 1985; Sternberg & Powell, 1983), Bonacci (1993) showed that less-skilled readers ignore important and explicit context information that could be used to determine the meaning of unfamiliar words. Even within grades this phenomenon occurs. Good readers seem to know how to deal with unfamiliar words in context, while poor readers experience more difficulties. With regard to the studies used in the meta-analysis, poor readers might not only be unfamiliar with the target words, but also not understand other words in the text, which would perhaps obstruct them completely. Good readers would benefit from knowing a large number of words, which could give them more information about the unknown target words (Shefelbine, 1990). On the other hand, one could argue that good readers would know more words leaving less words in any given text to learn. However, the target words used in the studies were selected in such a way that they were mostly unknown, thereby giving everyone a chance of learning their meaning. Thus, we expect good readers to gain more words incidentally. Level of reading ability was coded in three categories: high ability readers, average ability readers, and low ability readers. Since we had to accept the level indicated by the researcher(s) in our coding, it has perhaps become a somewhat crude factor.

The third group of factors involves assessment variables. Response selection was coded to account for differences due to the use of multiple-choice tests or open-ended questions. Multiple-choice tests perhaps could be more sensitive in detecting small increments in word knowledge than tests with a definition format, in which subjects have to come up with a word meaning out of nothingness. We know that word learning is a gradual process in which subjects expand their knowledge of the word after each new encounter with the word (Jenkins et al., 1984; Nagy et al., 1985). We however can not be sure about the effects of the test format. Multiple-choice tests might be very suitable in cases where subjects encounter words for the first time and acquire only some general knowledge of the word rather than an increase in specific knowledge. If however any encounter with words moves them to a slightly higher level of word knowledge, open-ended tests might be as sensitive as multiple-choice tests. Response selection was coded yes for multiple-choice tests; no for open-ended questions.

Partial word knowledge was coded to differentiate between studies giving credit for imperfect word knowledge, and those only fully approving or rejecting an answer. Since we see word learning is a gradual process, we cannot expect subjects to give a complete and accurate description of the word meaning. Instead, we expect that the average subject will only have a general idea of what a difficult word means. It is therefore likely that studies using tests that account for partial word knowledge yield higher effect sizes than studies accepting only a right or wrong answer. Partial word knowledge was coded as no credit versus credit only on posttest versus credit on both pretest and posttest.

A final group of factors relates to the material and the task. Reading purposes supposedly provoke subjects to proceed with reading tasks in different ways.

The reading purpose is the task that subjects receive before reading the text(s) in which the unknown words appear. Reading purpose has been shown to influence the way in which a text is processed (Klauer, 1984). Incidental word learning will be more successful with more rather than less elaborate processing (Hulstijn, 1999). The attention of readers will be affected by the goal they have in mind (Barnes, 1986), which in turn influences comprehension and recall (Beck, Perfetti, & McKeown, 1982). Thus, the way in which the text is processed will influence the amount of attention given to unknown words and the retention of the word meanings. We expect groups that read without specific purpose to learn less word meanings incidentally than groups that know that questions about the text(s) will follow. We coded this factor as read the text versus study the text.

The coding of the authenticity of the text reflects the origin of the text: Did the researchers use an existing text or did they write a text for specific research purposes? If Nagy et al.'s (1985) statement that most normal texts give only little information about word meanings as opposed to unnatural text is true, then our meta-analysis should show that studies using specially constructed texts yield larger effect sizes than other studies. Authenticity was coded as specially constructed versus not specially constructed.

The last factor that was coded was the text-target word ratio. This represents the average amount of text surrounding each target word. A short text that holds lots of unknown words will not give the reader much contextual aid, while a long text with only a few unknown words will give more opportunity to derive, at least, the meaning successfully. This, of course, is a necessary condition for incidental word learning. Browne (1989) showed that college-age subjects reading texts with a high proportion of infrequent words provided fewer correct English synonyms for nonsense words than did subjects reading texts with a low proportion of infrequent words. Therefore, we expect more incidental word learning when there is a large amount of surrounding text, which means that the text-target word ratio is relatively high. The ratio is derived from the number of target words in the experimental passages and length of the passages measured in words. Text length and number of target words were both coded in an open-ended way.

Interrater Reliability

Three coders independently coded all studies on all factors. In order to measure the agreement between the evaluations of the different raters, Cohen's kappa was calculated for nominal variables, and r_1 (design 2) for continuous ones (Orwin, 1994). Furthermore, an overall rate of agreement was calculated for all variables. Differences in coding were resolved by a majority decision rule.

After an initial phase of coding, it was decided in mutual agreement that some adaptations to the coding scheme had to be made. The categories of four variables were reduced, altered, or made more explicit. A new phase of coding the four variables resulted in the interrater-reliability figures as stated in Table 2.

Description of the Studies on Incidental Word Learning

The total number of subjects in the 20 experiments is 2130, in experimental

TABLE 2
Reliability of coding by study characteristic: kappa/r_1 and agreement rate in percentages (AR)

Study characteristic	K[a]	r_1[b]	AR [c]
Study design	1	-	100
Pretest sensitization	.71	-	75
Time pretest-reading	-	.99	75
Time reading-posttest	-	.90	80
Grade	-	1	100
Level of reading ability	.93	-	90
Response selection	.76	-	85
Partial word knowledge	.71	-	75
Reading purpose	.78	-	85
Authenticity	.78	-	90
Text length*	-	.89	70
Number of target words*	-	.95	75

[a] = generalized Cohen's Kappa for nominal variables
[b] = generalized intraclass correlation (design 2) for continuous variables (Orwin, 1994)
[c] = percentage of observations agreed upon by the three coders
* text length and number of target words together = Text-target word ratio

as well as in control groups. In general, studies use an untreated control group design with pretest or a design in which two groups of students read only one text, but are tested on words from both the text they read and the text the other group read. This second design, sometimes accompanied by a pretest, was used in 8 out of the 20 experiments. This factor is not the only one that varies across studies: Sample sizes vary widely (19-392), as do all other coded variables. This does not mean that the studies cannot be compared. Glass et al. (1981), Schwarzer (1989), and Wolf (1986) all state that it is in fact the advantage of a meta-analysis to enable us to compare different studies, provided they measure the same underlying construct.

Obviously, pretest sensitization was only a threat to the validity of studies using a pretest. More than half of these studies used pretests containing distractor items; seven experiments tested only the target word knowledge. Six of these seven studies assessed knowledge of these words in isolation only 1-3 days before the actual reading of the text. It is remarkable that the studies that pretest only the target words tend to stay close to the reading experiment. In the majority of the studies where the pretest contains distractor items the time between pretesting and reading is 7-29 days. Knowledge of the target words was mostly tested immediately after the reading of the text(s). Only one study (Nagy et al., 1987) delayed testing until a week after reading the texts.

As to the subject factors, a large range of grades is represented in this meta-analysis, namely from 3rd grade to 11th grade. The average grade in the studies was the 6th. The level of reading ability was also subject to variation: Some studies used students from the whole ability range, but mostly only average and above average students participated.

Regarding the assessment factors, only five studies (6 effect sizes) used real multiple-choice tests for a first assessment of word knowledge after reading the

text. In other studies students were asked to provide definitions or answer yes/no questions. In one study (Nagy et al., 1985) students were interviewed on the meanings of the target words.

Only two studies (Nagy et al., 1987, Stahl, 1989) did not account for partial word knowledge. Other studies assessed partial word knowledge in one of two ways: on the one hand, studies using the definition format gave credit for only partial word meanings. On the other hand, studies using multiple-choice tests or yes/no questions always had more than one item per word, and the different items per word differentiated between only a global notion of what a word means and more profound word knowledge. Pertaining to the definition format, the refinement of scoring schemes varied from one study to another. For example, Konopak (1988b) uses a 0-5 point scale, while Granick (1997) used a 0-2 point scale, and Nagy et al. (1985) used a 0-3 point scale.

The material-related factor reading purpose was different for all studies. Some researchers told their subjects that they were interested in finding out how children learn from reading, others told their subjects that they would have to answer topic questions afterwards, still others just told the students to read the text. The broad distinction that could be made between the studies was whether or not the subjects knew that questions were to follow the reading or not.

The text-target word ratio ranged from 37 words for every target word (Diakidoy, 1993, exp.1) to 115 words per target (Kranzer, 1988). Most studies used existing texts; only 3 experiments used newly written ones (Diakidoy, 1993, exp.1 & 2; Durkin, 1990). Diakidoy used two texts in her experiments: One text comprised a large number of clues, the other one comprised hardly any clues. Incidental word knowledge gain was assessed as an average of these conditions, which we interpreted as an assessment of incidental word learning under natural reading circumstances. Some of the existing texts used in the other studies were shortened in length or adapted to the age group.

Results of the Meta-Analysis

As a basis for analysis, a random effects model was chosen, since it can account for large variation between studies. We used the computer program VKHLM (Bryk, Raudenbusch & Congdon, 1994) to estimate the random effects model.

Effect sizes ranged from logit (p) = -3.55 (Gordon et al., 1992, condition A) up to logit (p) = .16 (Konopak, 1988a). A combined effect size logit (p)' for all studies of -1.70 (se = .21, p = .00) was computed, taking into account differences in sample size. This effect size was transformed back to an average probability of incidental word learning of .15. A corresponding 95% confidence interval of .11 to .22 was obtained. A test of homogeneity indicated that the results were heterogeneous (X^2 = 121.33, p = .00). The percentage of variance in scores that can be explained by sampling error is 81. The remaining variance (19%) has to be explained by systematic factors. This may be hard, as Hunter, Schmidt, & Jackson (1982) give as a rule of thumb that less than 25% of remaining systematic variance may be negligible in magnitude. It could be hard to find moderator variables. However, since the chi-square test of homogeneity is significant, thus indicating substantial heterogeneity, we decided to further explore, in a conservative manner, the variation in effects.

TABLE 3
Summary of features of the studies included in the meta-analysis

Study	*N*	PPD	PS	Time 1	Time 2	Grade	RA	RS	PWK	RP	A	Ratio
Diakidoy 1993 exp.1	50	1	1	14	0	6	HAL	0	2	Read	Na	37
Diakidoy 1993 exp.2	73	1	1	14	0	6	HAL	0	2	Read	Na	44
Durkin 1990	54	1	1	29	0	5	HAL	0	1	Study	Na	87
Gordon et al. 1992 H	19	1	0	1	0	5	H	0	2	Study	A	79
Gordon et al. 1992 A	19	1	0	1	0	5	A	0	2	Study	A	79
Granick 1997	349	0	-	-	0	8	HAL	0	1	Read	A	102
Herman 1985	39	1	1	14	0.5	8	L	1	1	Study	A	43
Konopak 1988a H	27	1	0	3	0.51	11	H	0	2	Study	A	150
Konopak 1988a A	28	1	1	3	0.5	11	A	0	2	Study	A	150
Konopak 1988b H	27	1	0	3	0.5	8	H	0	2	Study	A	100
Konopak 1988b A	25	1	0	3	0.5	8	A	0	2	Study	A	100
Konopak et al. 1987	21	1	1	0	1	11	HAL	0	2	Read	A	150
Kranzer 1988	19	0	-	-	0	8	HAL	0	1	Study	A	115
Nagy et al. 1985	57	1	1	3	0	8	H	0	1	Study	A	66
Nagy et al. 1987	352	1	1	14	7	5	HAL	1	0	Read	A	43
Schwanenflugel 97	33	1	1	7	3	4	HAL	1	2	Study	A	99
Shu et al. 1995 exp.1	146	1	1	7	0	4	HAL	1	1	Read	A	42
Shu et al. 1995 exp.2	301	1	1	7	0	4	HAL	1	1	Read	A	75
Stahl 1989	91	0	-	-	0	6	HAL	1	0	Study	A	50
Stallman 1991	100	1	0	21	0	5	HAL	0	2	Read	A	38

N: number of students; PPD: 1 = pretest posttest design, 0 = other design; PS: 1 = presence of distracter items on pretest, 0 = absence of distracter items on pretest, - = no pretest; Time1: time in days between pretest and reading; Time2: time in days between reading and posttest; Grade: average grade of students in study; RA: H = high ability readers, A = average ability readers, L = low ability readers; RS: 1 = multiple choice tests or yes/no tasks, 0 = open-ended questions or interview; PWK: 2 = credit for partial word knowledge on both pretest and posttest, 1 = credit for partial word knowledge on posttest only, 0 = no credit for partial word knowledge; RP: read = read the text, study = study the text; A: Na = texts especially constructed for the experiment; A = authentic texts; Ratio = number of words for each target word.

TABLE 4
Results of the random effects model

Fixed effect	Coefficient	*se*	*T*	*p*
	p = -1.7036	.21	-8.214	.00
Random effect	Population variance	*df*	X^2	*p*
	.63	19	121.33	.00

Explorative Investigation of the Variation in Effects

In order to explore the variation in effects, a multi-level regression analysis was performed to identify certain study characteristics that explain the heterogeneity of results (Hox & De Leeuw, 1997). Analyses were performed with the program VKHLM. The following hierarchical model is used:

$$\text{logit}\,(P_{ij}) = \gamma_0 + \gamma_1 W_{1j} + \gamma_2 W_{2j} + \ldots + \gamma_s W_{sj} + u_j$$

in which logit (p_{ij}) is the mean effect size, $\gamma_0 \ldots \gamma_s$ are regression coefficients, $W_{1j} \ldots W_{sj}$ are study characteristics, and u_j is a random error (Bryk & Raudenbusch, 1992).

First of all, for purposes of exploratory analysis, study characteristics were separately entered into the regression equation in order to identify possible relevant predictors, which might not surface once the conservative model building phase began. This is an important step, because of the relatively small number of effect sizes and the resulting limited power of analysis.

In the second phase of analysis we combined separate predictors in one model in order to account for the systematic variation in the effect sizes. Predictors were entered into the model according to the following strategy: first, study conditions were entered, to remove variance due to design features; then the subject factors, to remove variance due to student characteristics; then the assessment factors to control for test-related variance; and finally the text and word factors for a conservative estimation of their impact. Initially significant predictors of each group were entered one-by-one into the equation in order to retain power. Significant predictors of each group were retained in the regression equation when entering subsequent groups. This way a conservative manner of identifying factors explaining additional variance was ensured.

In the initial analysis of the impact of separate study characteristics, several predictors reduce the variance in effect sizes significantly. The study conditions group—design, pretest sensitization, time between pretest and reading, and time between reading and posttest—yielded one significant predictor, namely pretest sensitization. Studies using pretests without any distractor items yield larger effect sizes than studies where distractor items on the pretest are present. Pretest sensitization was significant at the p = .04 level, explaining 22 percent of the systematic variance in the effect sizes. If all other variables are held stable, a

pretest with distractor items will give a word learning probability .11; a pretest without any distractor items will raise this probability to .23.

Both subject factors—grade level and reading ability—proved to be significant in the initial analysis. Grade level explained 46 percent of the systematic variance in effect sizes ($p = .00$); level of reading ability 43 percent ($p = .01$). Everything being equal, Grade 4 subjects will show a probability of learning a word of .08; Grade 11 will show .33. As to reading ability, low ability readers will gain about .075, average ability readers .12 , and high ability readers .19.

The assessment factor response selection just failed to be a significant predictor ($p = .057$), but results were in the expected direction, favoring multiple choice tests over open-ended tasks. Partial word knowledge, the other assessment variable, did reduce systematic variance in effect sizes significantly, by 47 percent ($p = .00$). Studies giving credit for partial word knowledge yield higher effect sizes than studies that do not take partial word knowledge into account. If no credit is given for partial word knowledge the probability of learning a word will be about .06; if credit is given only on the posttest, the probability will be .12; credit on both pretest and posttest will increase the probability to .23.

The text and word factors—reading purpose, authenticity, and text-target word ratio—yielded one predictor that explained variance, namely the text-target word ratio. A low density of unknown words in a text produces a higher word learning chance than a high density of unknown words. The text-target word ratio reduces systematic effect size variance with 32 percent ($p = .01$). If the density of unknown words in a text is low, for example 1 word on every 150 words, the probability of learning a word will be about .30, 1 unknown word on every 75 will yield .14, 1 unknown word on every ten words gives .07.

The predictors that were significant in the initial phase of analysis were used for further model building. Modeling started with the factor pretest sensitization. This factor was no longer significant when the two subject factors (grade level and level of reading ability) were entered into the equation. Pretest sensitization was dropped from further model building analysis, as was level of reading ability. Level of reading ability was equally no longer significant in combination with grade level. Partial word knowledge was added to the regression equation that contained grade level. Both predictors stayed significant and were maintained for further analysis. Text-target word ratio was entered last. This predictor did not add significantly to the variance that was already explained by grade level and partial word knowledge.

Together, grade level and partial word knowledge explain 66 percent of the systematic variance in the effect sizes. Grade level by itself reduced true parameter variance from .63 to .34; a reduction of variance by 46 percent. Partial word knowledge additionally reduces true parameter variance to .21, a further reduction by 20 percent.

In the end, although grade level and partial word knowledge taken together explain 66 percent of the systematic variance, results remained significantly heterogeneous: Not all systematic variance is explained ($X^2 = 59.93$, $p = .00$).

Sensitivity Analysis

Some possible threats to the results of the meta-analysis were evaluated in further analyses. To find out whether some studies influenced the results too

TABLE 5
Results of the effect size analysis

Fixed effect	Coefficient	*se*	*T*	*p*	Cumulative variance explained
Base	-3.71	.45	-8.21	.00	
Grade level	.18	.07	2.79	.01	46%
Partial word knowledge	.56	.21	2.67	.02	66%
Random effect	**Population variance**		***df***	X^2	***p***
	.21		17	59.93	.00

much, studies were left out one by one. The mean effect size logit (p)' ranged this way from -1.82 to -1.70 (-1.70 with all studies included). The 95% confidence interval was originally -2.11 to -1.29; leaving one study out repeatedly widened the range from -2.19 to -1.29. Since these figures stay close to the original values, it is not likely that one study or another influences too strongly the overall outcome.

A boxplot of sample size versus probability indicated Konopak's study (1988a; both experiments) to be an outlier. However, inspection of the study did not give us any reason of leaving this one out. No abnormalities were found. Also, new multi-level regression analyses leaving out the studies with the lowest and the highest effect sizes, indicated the same predictors to be significantly explaining the variance in results.

Discussion

Incidental word learning during reading has been assumed to explain the large growth in vocabulary knowledge for a long time (Beck & McKeown, 1991; Hafner, 1932). Evidence for this assumption came only in 1985, with a study by Nagy, Herman, and Anderson. In a series of subsequent studies these and other researchers found word learning chances of varying magnitude. The available studies differ in many respects: a variety of designs, student populations, assessment instruments, and text and word materials is involved. The statistical meta-analysis we performed has confirmed that incidental word learning during natural reading takes place and has explained a large part of the variation in outcomes between studies.

We found a mean effect size of logit (p) = -1.70. This effect size has been derived from probability estimates and can be translated back into a mean probability of learning an unknown word while reading of .15. The results of the studies we analyzed turned out to be heterogeneous. This means that there have to be some underlying factors that explain the variation in outcomes between studies. In order to identify these conditions we performed an explorative hierarchical multi-level regression analysis.

Initial exploratory analyses of the impact of single factors indicates that pretest sensitization, grade level, level of reading ability, partial word knowledge, and text-target word ratio are significant predictors of the probability of learn-

ing a word incidentally. Results are in the expected direction: Absence of distractor items on the pretest results in a higher probability estimate; the higher the grade of the students, the more words they learn incidentally; the higher their reading ability, the more they learn; studies that use assessment methods sensitive to partial word learning show higher word learning gains; students learn more words when the ratio of text to target words is higher.

In a more conservative analysis, a combined model of the significant predictors shows that students' grade level and partial word knowledge together explain a large part of the variation in effect sizes between studies (66%).

Implications for Future Research

Our meta-analysis clearly indicates that students learn words incidentally while reading. Future research is needed on the factors that influence incidental word learning across studies. In an exploratory way we identified five factors. Part of the variance, however, remains to be explained. Questions remain as to what the influence of design characteristics is, and whether non-significant predictors at the text and word level should be left out in future research.

Our explorative analysis indicates that pretest sensitization, grade, level of reading ability, assessment methods sensitive to partial word knowledge, and text-target word ratio influence the probability of word learning while reading. Pretest sensitization should be taken into account in future research. If researchers decide to administer a pretest, distractor items should be included. Further, a separate study of effects of pretesting seems worthwhile. The different Konopak studies (1988a, 1988b) demonstrate relatively high probabilities of learning a word, which might be caused by the absence of distractor items on the pretest. In addition, these pretests were administered shortly before the actual reading of the texts, which makes it all the more probable that the attention of the students was directed towards the target words.

In our analysis, the subject factors grade and reading ability are interrelated, since reading ability was not significant when entered together with grade level into the regression equation. More research needs to be done on this relationship and on their separate effects. Older and more able students perhaps learn more word meanings by using some strategies during reading which help them derive and learn words. It would be worthwhile to examine which strategies these are, and see if they can be taught to others. It is further important to assess how many words are learned by children of different ages, while reading age-appropriate materials.

The outcome of reading ability as a significant predictor is very interesting, since not all studies in the meta-analysis demonstrated this effect. Nagy et al. (1985) were not able to demonstrate a significant effect of reading ability; in this study there was only a slight indication that high ability readers learned more. A significant effect of reading ability was also absent in the studies of Nagy et al. (1987) and Shu et al. (1995). In the studies of Herman (1985) and Konopak (1988a, 1988b), however, the influence of this predictor was significantly present. The lack of significant findings in some studies indicates the necessity of using a large population of subjects in subsequent studies, in order to have sufficient statistical power.

Only a few studies (Nagy et al. 1987; Shu et al., 1995) used multiple grades.

Although gains were found for all grades, these studies failed to provide us with the statistical data necessary to distinguish between the grades. In this meta-analysis, grades had to be averaged, at the expense of the sensitivity of the analysis. In spite of this, our meta-analysis demonstrates higher grades learn significantly more words. In order to confirm these findings, future studies should include distinct grades and report separate figures for each age group.

Studies using assessment methods sensitive to partial word knowledge show higher amounts of incidental word learning. This finding confirms the incremental nature of word learning. One cannot expect students to learn a dictionary-like meaning of a word after only one encounter. In order to account for even the smallest gains in word knowledge, studies should use assessment methods sensitive to small increments in word knowledge. It is likely that the lack of credit for partial word knowledge in Nagy et al.'s 1987 study is one of the factors that explains the divergence in probability of learning a word from their 1985 study (.05 vs. .11).

Pertaining to the text-target word ratio, it would be interesting to include in future studies a measure of the students' existing word knowledge, since a large vocabulary facilitates learning other words. If there is only one unknown word in a large body of text, it seems easier to learn the word's meaning than when there are more unknown words in the surrounding text. Since this familiarity with the text varies from individual to individual, an assessment of word knowledge of the surrounding text is warranted.

Beyond pretest sensitization, grade level, level of reading ability, partial word knowledge, and ratio of unknown words, no other predictors were significant. However, we do not feel we can dismiss all other study characteristics as irrelevant. The design, subject, assessment, and material factors in our analyses are the ones that are relevant to all (or almost all) of the studies we collected. Within separate experiments these factors are generally not manipulated. In future research several of them deserve more attention.

For example, more research is needed on the effects of time intervals. An important part of incidental word learning is the retention phase. What is remembered of a word meaning a week or even a month after reading? In our meta-analysis there was only one study assessing incidental word learning one week after reading. This study (Nagy et al., 1987) was found at the lower bound of the confidence interval, reporting a five percent chance. Unfortunately, one study is not enough to explain significant variation between studies. For future research, one should vary the moment of assessment. Research on second language reading has indicated that deriving word meanings successfully does not automatically lead to retention a few days after the reading (Mondria & Wit-De Boer, 1989). One could imagine that the degree of retention is even less after a longer time interval. On the other hand, in a longer time span, subjects might encounter the same word again which would again expand their knowledge of the word. For an investigation of these issues, nonsense words could perhaps be used. The effects of the use of nonsense words to measure incidental word learning would then have to be compared to the use of existing words.

More research also needs to be done on the way in which incidental word learning is assessed. Although no differences were found between multiple choice measures and open-ended tasks, a further investigation of types of measurement

instruments seems warranted. A combined investigation of varying retention time intervals and effects of recognition versus production measures may yield interesting results, such as increasing differences on productive tasks as opposed to receptive tasks when there is an increase of the time interval between reading and testing.

Reading purpose is a factor that has to be further explored too, simply because children read a variety of texts with different purposes and encounter many words multiple times. Reading purpose was roughly translated into read or study the text. However, children read for a variety of reasons, in and out of school, and it would therefore be interesting to look at the different reading tasks and examine what each means for incidental word learning. One can imagine that reading comprehension tasks that are assigned to students in school provoke very different behavior among children than does free reading, while both can be considered as being part of natural reading.

Future research should also focus on finding some new predictors. Some researchers already included variables such as cultural and linguistic background of the students, word frequency, morphological transparency of the target words, and their conceptual difficulty. One factor that has to be and can be accounted for in future experiments is part of speech. One might expect the meaning of nouns to be easier to derive than that of adverbs, since nouns are often more concrete and imaginable. However, Schwanenflugel et al. (1997) have found some surprising evidence favoring the acquisition of non-nouns over nouns.

Most studies in the meta-analysis pertain to the English language. In one study (Shu et al., 1995), native Chinese students learning Chinese were tested. Results indicate that Chinese students also learn new word meanings incidentally, which might indicate that learning words from context is a universal phenomenon. Involving other languages could yield interesting results.

Instructional Implications

Educators will be interested to know that the average probability to learn unknown words from context is about 15 percent. Under natural reading circumstances students will spontaneously derive and learn the meaning of about 15 words on every 100 unknown words they encounter.

Incidental word learning is a process, which consists of deriving the meaning of an unknown word and storing that meaning. It depends on several factors whether this process is successful. In order to arrive at learning new word meanings incidentally, students will have to be able to derive word meanings from context. A recent meta-analysis (Fukkink & De Glopper, 1998) and a recent review (Kuhn & Stahl, 1998) of the effects of instruction in context use on deriving word meanings show that students can learn to become skilled in deriving word meanings. Since knowing how to derive word meanings from context is a necessary condition for incidental word learning, practice or instruction in this skill would be expected to transfer to the incidental word learning ability.

Remembering the meaning of a word after it has been derived is the second part of the process. The skill in this process will also vary between subjects. Some students will have more trouble than others remembering what they have learned. It seems plausible that incidental word learning ability can be fostered

by training students' memory skills. One successful method for practicing memory skills is the keyword method of Pressley, Levin, and McDaniel (1987). Perhaps a combined training of word deriving skills and memory skills may have an impact on incidental word learning.

Subject factors that influence the incidental word learning chance seem to be the students' grade and level of reading ability. Students of high and average reading ability will learn more than those of low ability. Also, older students will learn more than younger ones. These findings are probably interrelated; reading ability was no longer a significant predictor when grade level was entered into the regression equation. It might well be that with age and level of reading ability the use of certain reading strategies increases. Perhaps not only younger readers, but also readers of low ability, lack the strategies to derive the meaning of an unknown word or have trouble remembering what they have derived.

Our analysis provides support for the idea that one can get better at incidental word learning. Very young children learn words from context, thus suggesting incidental word learning is an innate skill that might not be amenable to instruction. Our results, however, suggest that children improve the skill to handle unknown words over time. Subjects from Grade 4 already demonstrate some knowledge as to deriving and remembering word meanings. By the time they are in Grade 11, if all other variables are held constant, the probability that a word meaning is learned from context is four times as high. If we can grasp in what respect students change over the years we might be able to help and teach others—for example, children with reading deficiencies. Evidence for growth in this aspect comes from cross-sectional studies of derivation skills that show an increase in performance with age (Carnine et al., 1984; Werner & Kaplan, 1952). It should be further investigated which deriving skills and memory skills play a role in the age-related development of the incidental word learning ability.

Students' vocabulary size seems to be another subject factor of major importance. Due to the lack of information in the studies we analyzed, students' vocabulary size was not part of the set of predictors in the meta-analysis. Shefelbine (1990) states that low vocabulary students are at a disadvantage when it comes to independently acquiring new vocabulary, because they encounter too many unknown words when they read. Our explorative multilevel regression analysis showed that a high density of unknown words in a text obstructs incidental word learning. If students only have a small vocabulary, it is therefore advisable not to read texts above grade-level, in which too many unknown words figure. Also, reading those texts is likely to discourage the low vocabulary students, which according to Shefelbine would further compound the vicious circle of the poor who get poorer relative to the rich who get richer.

On the other hand, Carver (1994) indicates that students encounter only few unknown words in texts at their ability level, which prohibits students from learning new word meanings. Students should therefore read slightly above grade-level. This is of course an area of tension. In order not to hinder comprehension, students must not encounter too many unknown words; in order to learn new word meanings students have to encounter unknown words. Depending on the purpose of reading—a reading lesson or reading for a specific subject matter like math or biology—and depending on the reading ability of the stu-

dent, teachers should ensure that the density of unknown words in the material they offer is suitable for the individual student.

Other factors influencing incidental word learning are probably the purpose by which students read a text and the material they read. Although the reading purpose was not a significant predictor according to our explorative regression analysis, one can imagine that reading out of school for fun does not yield the same results as a reading comprehension task in school. Different factors are at stake here, relating to the depth of text processing and the amount of attention given to the words (Klauer, 1984). These factors cannot be easily influenced in a way that more words would be learned. One could of course gloss the meaning of difficult words in the margin of the text. However, the question then would be whether such a text manipulation would still allow for incidental word learning. Further research would have to indicate what the influence of this manipulation would be, both on the word learning chance and the student's purpose and motivation for reading the text at hand.

A final, interesting issue for educators pertains to the extent that incidental word learning contributes to students' vocabulary growth. From our estimate of the probability of learning unknown words from context, it will be clear that natural reading has the potential to make contributions to vocabulary growth. Questions remain as to the actual contribution reading makes. In order to estimate the absolute or relative contribution of reading to vocabulary growth, information is needed on several parameters.

To begin with, students' reading volume needs to be estimated. How many words do students annually encounter while reading? Nagy et al. (1987) present estimates of the volume of free reading students do, but as yet these estimates lack empirical verification. Next to free reading, the reading that students do in school should be included in the estimation of their reading volume, insofar as this reading is not accompanied by vocabulary instruction. Here, recent relevant data seem to be lacking.

Secondly, information is needed about the amount of unknown words students encounter. A study on this issue was carried out by Carver (1994). He concluded that the percentage of unknown words is approximately 1% when the difficulty of the reading material matches the ability level of the individual reader. When the material being read is relatively easy, close to 0% of the words will be unknown. Relatively difficult reading materials will contain around 2% unknown words. Carvers' data are based on students' self assessment of word knowledge. Students had to underline the unknown words they encountered in the texts presented to them. In the sample of students of Grade 3 to Grade 6, forty percent of the sample failed to underline words that no doubt were unknown to them: metacognitive, rauding, and edumetrically. At present, the validity of the above estimates can therefore be called into question. An independent replication with alternative measures of word familiarity would therefore be highly desirable.

Precise and valid data on reading volume and frequency of unknown words are sufficient for estimation of absolute vocabulary growth from reading. Appraisal of the relative contribution of reading is only possible when a third parameter is known. How many words do students learn on an annual basis? Estimates of vocabulary growth vary widely (Beck & McKeown, 1991). Beck

and McKeown state that the figure of seven words per day seems to be reasonably well supported: Miller's (1991) and Aitchinson's (1994) estimate of daily growth is more than ten words. Anglin (1993) estimates that the amount of psychologically basic words learned per day actually increases from 3.3 (1.5 years – Grade 1) to 6.6 (Grade 1 – Grade 3) up to 12.2 (Grade 3 – Grade 5). What becomes clear from this variation in growth estimates is that we lack the data for an adequate assessment of the relative contribution of natural reading to vocabulary growth. What we do know however, from our meta-analysis, is that students have a fair chance of learning unknown words from reading. Natural reading has the potential to make a contribution to vocabulary growth.

References

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

Aitchison, J. (1994). *Words in the mind: an introduction to the mental lexicon* (2nd ed.). Cambridge, MA: Blackwell.

Anglin, J.M. (1993). Vocabulary growth: a morphological analysis. *Monographs of the Society for Research in Child Development, 58* (10, Serial No. 238).

Barnes, J.A. (1986). Reading comprehension: The role of schema 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.E. David Pearson (Eds.), *Handbook of Reading Research II* (pp. 789-814). New York: Longman.

Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506-521.

Bonacci, F.J. (1993). The acquisition of word knowledge from expository context in skilled and less-skilled readers. *Dissertation Abstracts International, 54*(05A). (University Microfilms No. AAI93-26302).

Browne, B. (1989). Effects of vocabulary difficulty on word definition and prose recall. *Journal of General Psychology, 116* (4), 385-392.

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

Caracciolo, C. E., Fabio, R., & Trombetta, C. (1988). Apprendimento incidentale e rievocazione in un disegno di transfert part-whole. Ricerca sperimentale in eta evolutiva [Incidental learning and recall in a part-whole transfer pattern: Empirical research with developmental age subjects]. *Studi di Psicologia dell'Educazione, 7*(2-3), 89-109.

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

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

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

*Diakidoy, I.-A.N. (1993). The role of reading comprehension and local context characteristics in word meaning acquisition during reading. *Dissertation Abstracts International, 54*(05A). (University Microfilms No. AAI9329013).

Dupuy, B. & Krashen, S.D. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning, 4*(1-2), 55-63.

*Durkin, P.M. (1990). Vocabulary acquisition in context reconsidered: The effect of word type and exposure level on the learning of unknown words. *Dissertation Abstracts International, 51*(08A). (University Microfilms No. AAI9102613).

Elivian, J. (1938). Word perception and word meaning in silent reading in the intermediate grades. *Education for September,* 51-56.

Eller, R.G., Pappas, C.C., & Brown, E. (1988). The lexical development of kindergartners: Learning from written context. *Journal of Reading Behavior, 20*(1), 5-24.

Elley, W.B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 14*(2), 174-187.

Fischer, U. (1994). Learning words from context and dictionaries: An experimental comparison. *Applied Psycholinguistics, 15,* 551-574.

Fukkink, R.G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: a meta-analysis. *Review of Educational Research, 68*(4), 450-468.

Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage Publications.

*Gordon, J., Schumm, J.S., Coffland, C., & Doucette, M. (1992). Effects of inconsiderate versus considerate text on elementary students' vocabulary learning. *Reading Psychology, 13*(2), 157-169.

*Granick, J.S. (1997). Learning word meanings of low frequency words from context during natural reading. *Dissertation Abstracts International, 58*(03A). (University Microfilms No. AAI9723793).

Hafner, R. (1932). Casual learning of word meanings. *Journal of Educational Research, 25*(4-5), 267-277.

Hedges, L.V. (1994). Statistical Considerations. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp.29-38). New York: Russell Sage Foundation.

*Herman, P.A. (1985). Incidental learning of word meanings from written contexts that systematically vary text features. *Dissertation Abstracts International, 46*(11A). (University Microfilms No. AAI8600208).

Hox, J. J. (1995). *Applied multilevel analysis* (2nd ed.) . Amsterdam: TT-publikaties.

Hox, J.J., & De Leeuw, E.D. (1997). Meta-analyse via multiniveau-modellen [Meta-analysis by means of multi-level models]. *Methods & Statistics Series, 58.* Amsterdam: Faculteit der Pedagogische en Onderwijskundige Wetenschappen van de Universiteit van Amsterdam.

Huckin, T.N., & Zhendong, J. (1986). Inferring word-meaning from context: A study in second language acquisition. *Proceedings of the Eastern States Conference on Linguistics, 3,* 271-280.

Hulstijn, J.H. (1999). *Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity.* Manuscript in preparation. University of Amsterdam.

Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). *Meta-analysis. Cumulating research findings across studies.* Beverly Hills, CA: Sage Publications.

Jenkins, J.R., Stein, M.L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21*(4), 767-787.

Klauer, K.J. (1984). Intentional and incidental learning with instructional texts: A meta-analysis for 1970-1980. *American Educational Research Journal,*

21(2), 323-339.

*Konopak, B.C. (1988a). Effects of inconsiderate versus considerate text of secondary students' vocabulary learning. *Journal of Reading Behavior, 20*(1), 25-41.

*Konopak, B.C. (1988b). Eight graders' vocabulary learning from inconsiderate and considerate text. *Reading Research and Instruction, 27*(4), 1-14.

*Konopak, B.C., Sheard, C., Longman, D., Lyman, B., Slaton, E., Atkinson, R., & Thames, D. (1987). Incidental versus intentional word learning from context. *Reading Psychology, 8*(1), 7-21.

*Kranzer, K.J. (1988). A study of the effects of instruction on incidental word learning and on the ability to derive word meanings from context. *Dissertation Abstracts International, 50*(04A). (University Microfilms No. AAI8914330).

Kuhn, M.R., & Stahl, S.A. (1998). Teaching children to learn word meanings from context: A synthesis and some questions. *Journal of Literacy Research, 30*(1), 19-38.

Martin-Rehrmann, J.J. (1990). The effects of context on the spontaneous instantiation of general terms by third, fifth, and seventh graders. *Dissertation Abstracts International, 51*(3A). (University Microfilms No. AAI9012652).

Mayberry, M., Taylor, M., & O'Brien-Malone, A. (1995). Implicit learning: Sensitive to age but not IQ. *Australian Journal of Psychology, 47*(1), 8-17.

McKeown, M.G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*(4), 482-496.

Miller, G.A. (1991). *The science of words.* New York: Scientific American Library.

Mondria, J.-A., & Wit-De Boer, M. (1989). Woorden leren door raden? De invloed van context op het raden en onthouden van woorden [Learning words from guessing? The influence of context on guessing and retaining words]. *Levende Talen, 443,* 497-500.

*Nagy, W.E., Anderson, R.C., & Herman, P.A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237-270.

Nagy, W.E., & Herman, P.A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M.G. McKeown, & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, New Jersey/ London: Lawrence Erlbaum Associates.

*Nagy, W.E., Herman, P.A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253.

Orwin, R.G. (1987). A fail-safe N for effect sizes in meta-analysis. *Journal of Educational Statistics, 8,* 147-159.

Orwin, R.G. (1994). Evaluating coding decisions. In H. Cooper, & L.V.E. Hedges (Eds.), *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

Pressley, M., Levin, J.R., & McDaniel, M.A. (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches. In M.G. McKeown, & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 107-127). Hillsdale, NJ/ London: Lawrence Erlbaum Associates.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York: Russell Sage Foundation.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183-192.

Schatz, E.K., & Baldwin, R.S. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly, 21*(4), 439-453.

*Schwanenflugel, P.J., Stahl, S.A., & McFalls, E.L. (1997). Partial word knowl-

edge and vocabulary growth during reading comprehension. *Journal of Literacy Research, 29*(4), 531-553.
Schwarzer, R. (1989). *Meta-analysis programs. Program manual.* Berlin: Institut fur Psychologie, Freie Universität Berlin.
Sénéchal, M., & Cornell, E.M. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28*(4), 360-374.
Shefelbine, J.L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior, 22*(1), 71-97.
*Shu, H., Anderson, R. ., & Zhang, H. (1995). Incidental learning of word meanings while reading: A Chinese and American cross-cultural study. *Reading Research Quarterly, 30*(1), 76-95.
*Stahl, S.A. (1989). Task variations and prior knowledge in learning word meanings from context. In S. McCormick, & J.E. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction* (pp. 197-204). Chigaco: National Reading Conference.
*Stallman, A.C. (1991). Learning vocabulary from context: Effects of focusing attention on individual words during reading. *Dissertation Abstracts International, 52*(07A). (University Microfilms No. AAI9136743).
Stanley, P.D., & Ginther, D.W. (1991). The effects of purpose and frequency on vocabulary learning from written context of high and low ability reading comprehenders. *Reading Research and Instruction, 30*(4), 31-41.
Stein, M.L. (1989). The effects of multiple exposures on learning vocabulary from context. *Dissertation Abstracts International,* 49(09A).
Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown, & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ/London: Lawrence Erlbaum Associates.
Sternberg, R.J., & Powell, J.S. (1983). Comprehending verbal comprehension. *American Psychologist, 38*(8), 878-893.
Werner, H., & Kaplan, E. (1952). The acquisition of word meanings: A developmental study. *Monographs of the Society for Research in Child Development, 15, Serial no. 51*, 1.
Wolf, F.M. (1986). *Meta-analysis quantitative methods for research synthesis.* Newbury Park, CA: Sage Publications.

Authors

M.S.L. SWANBORN is a PhD student at the SCO-Kohnstamm Institute, Faculty of Behavioral Sciences, University of Amsterdam, Wibautstraat 4, 1091 GM Amsterdam, The Netherlands; Machteld@educ.uva.nl. Areas of specialization include vocabulary, reading, and linguistics.
K. DE GLOPPER is a professor at the SCO-Kohnstamm Institute, Faculty of Behavioral Sciences, University of Amsterdam, Wibautstraat 4, 1091 GM Amsterdam, The Netherlands. Areas of specialization include vocabulary, reading, and writing.

Received January 11, 1999
Revision received April 9, 1999
Accepted June 30, 1999