ELSEVIER

Learning and Instruction 15 (2005) 23–43

Learning and Instruction

www.elsevier.com/locate/learninstruc

# Deriving word meaning from written context: a process analysis

R.G. Fukkink*

*SCO-Kohnstamm Instituut, Universiteit van Amsterdam, Wibautstraat 4, 1091 GM Amsterdam, The Netherlands*

**Abstract**

The paper reports on a study among primary-school students of the process of deriving word meaning from written context in a first language. A sequential analysis of think-aloud protocols revealed that the students inferred one or more meanings, checked their inferences and then rejected or accepted them. These activities were performed in a highly flexible manner and their order varied. Four major sequences, which were not equally effective, showed up in a cluster analysis. The paper concludes by discussing implications for instruction and think-aloud research.


*Keywords:* Deriving word meaning from context; Reading strategy; Think-aloud methodology; Sequential analysis

## 1. Introduction

Primary-school students encounter many unfamiliar words while reading. A conservative estimate is that students in middle grades encounter each year some 16,000–24,000 totally new vocabulary items (Nagy, Anderson, & Herman, 1987) in approximately a million running words of text (Nagy, Herman, & Anderson, 1985). Children develop impressive vocabularies at a phenomenal rate in the primary school

* Tel.: +31 20 5251351; fax: +31 20 5251200.
*E-mail address:* r.g.fukkink@uva.nl

0959-4752/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.learninstruc.2004.12.002

period, and the default explanation is that most vocabulary is acquired by deriving word meaning from context (Beck & McKeown, 1991). According to Nagy and Anderson (1984), word meanings derived from a written context account for at least a third of the total vocabulary growth during the primary school period, which has been estimated at 3000 words per year (Beck & McKeown, 1991; Nagy & Scott, 2000).

Deriving the meaning of an unknown word from the written context is a complex and demanding task. First, the complexity of the word is a crucial factor. The unknown word may be related to a known concept, in which case it is a 'simple synonym' (Durkin, 1990). An important proportion of the new words that primary-school students encounter during reading, however, are 'conceptually challenging words' (Durkin), that is, words with new labels that refer to unfamiliar concepts. Also the concreteness of the unfamiliar concept is related to the complexity of deriving word meaning from context. The abstract word 'dilemma', for example, is conceptually challenging for primary-school students, although they may be familiar with the related concepts of 'difficult', 'choice' and 'two options'. Second, the complexity of deriving word meaning from context is, of course, also influenced by the context. Contexts are helpful only to a certain extent, and they do not reveal the full meaning of a word, even when explicit clues are present (e.g. a synonym or antonym clue). Some contexts are even misleading (Beck, McKeown, & McCaslin, 1983; Schatz & Baldwin, 1986). Readers therefore often glean only partial word knowledge from context (Swanborn & de Glopper, 1999) and include both correct and false attributes in their definitions (Fukkink, Blok, & de Glopper, 2001). Students experience additional problems in formulating a decontextualized word meaning that is abstracted from the original context. However, previous studies have also shown that both young students and students with a low verbal ability understand the task and are capable of performing meaning-derivation activities that were previously ascribed only to mature students of high ability. Although older students generate more and better hypotheses, younger and older primary-school students seem to work in a qualitatively similar way in deriving word meaning from single contexts (van Daalen-Kapteijns, Elshout-Mohr, & de Glopper, 2001; Werner & Kaplan, 1952).

Many authors have advocated instruction in deriving word meaning from context, considering its intrinsic complexity and also its potential (Baumann et al., 2002). Jenkins, Matlock, and Slocum (1989: 218) note that this type of instruction has a 'sound and persuasive rationale', because students encounter a large number of unknown words, and even a small improvement in the ability to infer the meaning of these words results in a sizable number of words learnt. An evaluation of experimental studies has shown mixed results: some were successful, while others established only small and non-significant effects or even negative effects (Fukkink & de Glopper, 1998; Kuhn & Stahl, 1998).

A central problem that educational designers face in this area is a lack of knowledge of how to help students to acquire an effective strategy. We do not yet know what activities untrained primary-school students perform in deriving word meaning from written context, and there is only scarce knowledge of how strategy instruction can fit in with these natural routines (Fukkink & de Glopper, 1998).

Related to this question is the issue of how to encourage students to actually use the instructed strategy in practice during their normal reading at school and in their leisure time. Research into the process of deriving word meaning from context that may redefine existing instruction in this skill is therefore needed (Fukkink & de Glopper; Harmon, 1999).

Research in this area should aim to find sensitive and detailed ways to examine the processes of word learning and how learners take advantage of contextual exposures (Blachowicz & Fisher, 2000). Unfortunately, the sequential nature of the collected data has been neglected in current think-aloud research on reading. Pressley and Afflerbach (1995) note in their review of reading-process research that they could not locate a study in which the sequential nature of the data had been taken into account. Acknowledging the sequential nature in the analysis of the protocols is important, because students do not perform separate, disconnected activities but follow a certain solution path to arrive at a final outcome. Furthermore, studies in other cognitive domains have shown that both the presence and the temporal order of activities are related to the quality of the final product (Breetvelt, van den Bergh, & Rijlaarsdam, 1994). Preservation of the order of activities in the analysis of protocols is also important from an instructional point of view. Prevailing instruction is aimed at providing students with a systematic approach that they should apply during their normal reading. Students are often taught a generic, systematic approach or a specific algorithm (Kuhn & Stahl, 1998), and instruction follows a sequential format (Harmon, 1999). Thus, both the natural processes of students and the instructed strategies that should tap them involve sequences.

The present study addressed the question what activities primary-school students employ in deriving word meaning from written context. In this, a sequential analysis of think-aloud protocols was conducted. This question was studied by focusing on conceptually challenging words in contexts that provided support but not explicit clues. This paper discusses the results in the context of instruction in deriving word meaning from context and research on cognitive processes.

## 2. Processes in deriving word meaning from context

In their review of thinking-aloud studies of reading, Pressley and Afflerbach (1995) distinguish four phases for word-related activities, namely (a) deciding whether to expend effort to determine the unknown word meaning; (b) paying greater attention to the word and its context (i.e. searching for clues); (c) inferring a word meaning with the use of the context and (d) subsequent evaluation of the generated word meaning.

### *2.1. Deciding whether to expend effort to determine the unknown word meaning*

In the first phase, readers can evaluate the importance of an unknown word to the overall meaning of the text before deciding whether to expend effort to determine the meaning of that unknown word. If they consider it unimportant, they can skip the word and continue reading; if they decide that it is important, they pay greater attention to the problematic word.

### 2.2. Searching for clues

The actual process of deriving word meaning from context starts with a search for clues. Three sources may be involved in this, namely internal clues in the word, external clues in the context and personal knowledge from long-term memory, which is automatically activated during reading (Waern, 1988). By gathering information from different sources, a reader selects constraints with which to restrict the word's meaning and to delineate the problem space (Boettcher, 1980; McKeown, 1985). The reader may decide to read ahead or to backtrack within or across sentence boundaries. He or she may also paraphrase small or large fragments of the context. Also determining the type of word, activation of schemata and summing up clues are context-related activities in this phase (Pressley & Afflerbach, 1995).

The search for clues is often limited. Students remain at the sentence level if they feel sufficient information is available, and will search for other information only if this information is considered inadequate (Goerss, Beck, & McKeown, 1999; Harmon, 1998). Harmon notes that readers seldom rely on personal knowledge from long-term memory. In sum, students' orientation is primarily text-bound and local, although additional information may be gathered.

A reader may, for example, encounter the morphologically complex word 'dilemma' in a certain context (see Appendix B). The word contains internal clues but is not transparent, as younger readers cannot be expected to know the Greek morphemes 'di' and 'lemma'. External clues, however, can be found in the surrounding context ('choosing' is a theme in the text – or, more specifically, a difficult choice between two options both of which are unappealing – as are the concepts of negative emotions, an argument and copying).

### 2.3. Inferring a word meaning

Based on the information gathered in the search phase, an attempt is made to infer the meaning of the word. This inference can take the shape of a determination of the type of word, although the reader may also come up with a meaning for the difficult word by giving a global synonym or a superordinate or subordinate. A reader may generate one or more meanings (Harmon, 1998). For example, in the case of 'dilemma', he or she may come up with the superordinate concept of 'problem' but may also infer the wrong answer, namely 'copying'.

### 2.4. Checking and evaluating

An inferred meaning is finally checked and evaluated. Any information that is gathered in the search phase may serve as a constraint, and the reader matches these contextual constraints with the features of the candidate concept (see McKeown, 1985). The check of an assumed word meaning leads to a positive ('accept') or negative evaluation ('reject'). If according to the reader the generated meaning does not fit the context, the process of deriving word meaning is repeated (a new hypothesis is generated); if however the meaning does fit, the process ends. The process of deriving word meaning from context may also be aborted ('give up') if the reader decides that no word meaning can be inferred that makes more sense in the context.

A common tactic applied by students is to substitute their tentative answer for the unknown word and to insert it in the original sentence of fragment thereof (Harmon, 1998, 1999). This check is therefore mainly text-bound and local, in parallel with the search phase. The evaluation phase for the word 'dilemma', for example, may consist of substituting the word 'problem' for the unknown word, thereby checking its grammatical and semantic fit. Superordinates fit by definition, because a subordinate concept ('dilemma') is always an example of the superordinate category ('problem'), and the reader may decide to stop the process.

### 2.5. A model of deriving word meaning from context

The working model of deriving word meaning depicted in Fig. 1 is a concise model that describes the process of deriving word meaning from external context. It is assumed that activities in the search phase involve an orientation towards clues. Also personal knowledge may be involved. However, previous think-aloud studies have shown that significant activities that involve an orientation towards the context (e.g. activation of schemata from long-term memory, the actual reading, the gathering of clues) do not often show up in the protocols as a result of automatization (Waern, 1988). Paraphrasing of the context, however, becomes perceptible in the protocols.

A reader subsequently infers a meaning, based on one or more information sources (referred to as 'infer'). The reader may, for example, hypothesize that 'dilemma' means 'problem'. After this, he or she may check the appropriateness of the hypothesis by, for example, replacing the word 'dilemma' with the word 'problem' in the sentence where the former occurs. This check leads to an evaluation with two possible outcomes: the meaning may be confirmed or rejected. A rejection is followed by the search for a new word meaning, and the process is repeated. It is possible that the student skips the search phase this time, and immediately infers a new meaning. Inferring meanings is considered a recursive process, that is, several

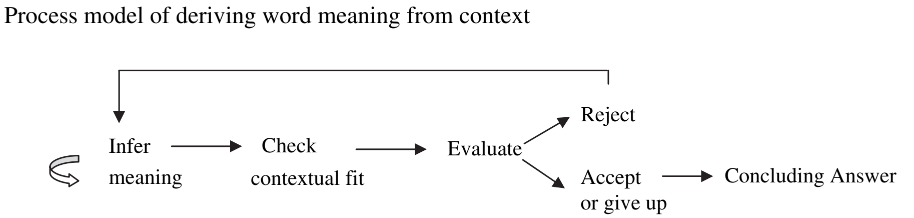


Fig. 1. Process model of deriving word meaning from context.

meanings may be formulated in succession. The process is concluded when the reader accepts the meaning as the final outcome (acceptance) or decides to give up.

The descriptive model not only postulates certain activities but also predicts the specific order indicated above. The following transitions are predicted: infer → check; check → accept; check → reject; accept → concluding answer; reject → $\text{infer}_{i+1}$ and $\text{infer}_i \rightarrow \text{infer}_{i+1}$. Furthermore, sequences with the order $\text{infer}_i$–check–accept–concluding $\text{answer}_i$ or $\text{infer}_i$–check–reject–$\text{infer}_{i+1}$ should be observed. Examples of each distinguished category as they appear in the protocols are given in the sequel.

The model is intended to capture the process of deriving word meaning as reflected in think-aloud protocols. Automatic processes are not verbalized in the think-aloud protocols (Ericsson & Simon, 1993), and have therefore not been included. Search activities, for example, are not likely to show up in think-aloud protocols, and activation of schemata is also considered an unconscious process, which does not leave traces in the protocols either (Waern, 1988). Other activities from the infer and check and evaluation phase are likely to show up in the protocols. Activities from the inferring phase show up as tentative answers. Checking and evaluation may also be identified. For example, a reader may read aloud a text fragment with the unknown word replaced by the students' tentative answer. The subsequent evaluation may also show up in the protocols in explicit rejections or confirmations (e.g. 'Yes, it is…' or 'No, it is not…').

## 3. Research questions

The first research question concerned the characteristics of the cognitive processes in primary-school students who are deriving word meaning from context, namely: what activities do primary-grade students perform in deriving word meaning from context? This question was addressed by looking at both separate activities and sequences. An explorative analysis was used to investigate relationships between these activities and their efficacy. The second question pertained to the relation between students' sequences and the quality of the derived word meanings at which they arrive: is a particular sequence of activities associated with better answers? It was hypothesized that the use of context in the evaluation of hypotheses is associated with better answers, because the semantic fit is explicitly checked.

## 4. Method

### *4.1. Participants*

The participants were students in grades 2, 4, and 6 at four different schools in Amsterdam. The students at these schools have varying cultural and socio-economic backgrounds. Ten students were randomly selected from each grade (total of 30 students), stratified for sex within grade. Students with teacher-reported reading difficulties were not included.

### 4.2. Materials

Following the designs of previous studies into deriving word meaning from context (van Daalen-Kapteijns & Elshout-Mohr, 1981; McKeown, 1985; Werner and Kaplan, 1952), identical texts were used to control word and text effects and to maximize the comparability of findings. First, target words were selected from a primary school dictionary (Verburg & Huijgen, 1994) to ensure that the concepts selected would be relevant, that is, representative of the words young readers encounter while reading. The sample of 12 words comprised six concrete and six abstract words. The average word frequency in the Celex database (*Center for Lexical Information*) was low (4.4 per million, ranging from 1 to 10 per million) to ensure that no words were used that students were already familiar with. All words were morphologically non-transparent (see Appendix A) to promote the derivation of word meaning from external context.

Short, narrative texts of approximately 100 words were subsequently constructed for each target word. The texts contained only high-frequency words and text difficulty was adjusted to an appropriate level for average readers at the end of grade 2 based on a reading difficulty index for primary-school students (Staphorsius & Verhelst, 1997); all students proved capable of reading the text. The texts contained no explicit clues (e.g. synonym, antonym or description clues). Target words were not placed in the first sentence of the texts (see Appendix B for some examples).

A version of the 12 texts with the target words deleted was presented to three adults, who were asked to fill in each cloze with an answer that was as specific as possible and fitted the context. Only four of the 36 answers did not match the concept of the deleted word (each exception concerned a different target word). The other answers, however, were identical to or synonymous with the deleted target word (58%) or closely related hyperonyms (31%) (e.g. 'to break' was filled in for the deleted target word 'to shatter'). For each target word, at least two of the three coders provided a correct answer. A subsequent check showed that all semantic elements could be retrieved in the aggregated set of protocols. For example, if three semantic attributes were distinguished for a word (A, B and C; see Section 4.5), all logically possible combinations were found in the protocols (i.e. not A, B or C; A; B; C; AB; AC; BC; and occasionally ABC), which indicated that there was at least some contextual support for each element. The texts were therefore considered to provide adequate contextual support.

### 4.3. Procedure

Each session started with a standardized instruction. Following the guidelines of Ericsson and Simon (1993), each student was instructed to 'talk aloud' at the beginning of the session. He or she was told:

> 'There's a difficult word in this text [experimenter points to a word printed in bold]. In this exercise, we'd like you to find out what this word means by

reading the text. I'm very curious to see how you do this and what you pay attention to. So keep talking aloud.'

This instruction specifically evoked activities related to deriving word meaning from context, and thus fitted in with the focus of the study. The procedure is also related to the classic distinction in the field between incidental word learning – which does not interrupt the natural reading process – and deliberately deriving word meaning from context, which does interrupt reading (Fukkink & de Glopper, 1998). Shefelbine (1990) characterizes incidentally and deliberately deriving word meaning from context as 'typical' and 'maximum performance', respectively. Whereas incidental word learning has not been a subject of think-aloud study, deliberately deriving word meaning from context has been studied with similar procedures and directions in think-aloud research on the same task in experimental studies (van Daalen-Kapteijns & Elshout-Mohr, 1981; van Daalen-Kapteijns et al., 2001; Waern, 1982, 1988; Werner & Kaplan, 1952) and in ethnographic studies (Harmon, 1998, 1999). The present study was thus focused on 'maximum performance' and on strategies that are used when students deliberately derive word meaning from context.

The students were told that they would be prompted to 'keep talking' if they were silent for some time. In order to separate the reading phase from the meaning derivation phase, they first read the texts and only then started to derive word meaning from the context. If thinking aloud stopped for some time after an answer had been formulated or if the student indicated that he or she had found the definitive answer, the student was asked: 'What definition do you think the dictionary would give for this word?' The answer to this question was the concluding answer (CA).

Trials were introduced that were similar to the experimental task. During these trials, confirmative feedback was given to encourage the subjects to continue talking aloud. The experimenter moved on to the experimental tasks without mentioning this to the student after an informal check had confirmed that the subject had talked aloud satisfactorily during three successive practice items, had performed the task as intended (i.e. had derived word meaning from context) without asking for clarifications, and had not produced social verbalizations. After this, feedback from the experimenter was confined to standard positive feedback after the completion of each task. The order of items was randomized for each subject. The sessions, each of which lasted 30–40 min, were tape-recorded and then transcribed for coding.

### 4.4. Coding of activities and quality of students' definitions

Protocols were first divided into utterances. A new utterance was started if a new grammatical subject or finite verb appeared. Sentences with the conjunction 'or' were identified and split into two utterances in order to separate semantically different answers (e.g. 'I think it's a small wall [1] or a fence [2]').

The coding scheme consisted of five main categories derived from the working process model (see Fig. 1). The listed examples were extracted from the protocols for the target words (see Appendix B).

Any utterance that contained a possible meaning of the target word was coded as an example of the Infer category (I; e.g. ‘A dilemma is perhaps disagreement’ and ‘A tough decision’). In many protocols, more than one word meaning was inferred. Semantically different hypotheses for one word were therefore indexed (e.g. the fragment ‘Being smart?…Erm, searching, well…If you find them being careful about it’ was coded as $I_1$, $I_2$, $I_3$). Similarly, each concluding answer was matched to a meaning with the same index ($CA_1$, $CA_2$, etc). If the concluding answer was semantically different from the hypotheses, a new index was added.

Check (Ch) was coded if an utterance was related to the check of a meaning. A meaning may be checked by focusing on the form of a word, its morphology, the context of the word or personal knowledge (for ‘dilemma’: ‘perhaps argument’ [$=I_1$] ‘…one says “I have, no you have done it” and the other says, er, the one who you suspect, she says “I didn’t do it”…one says “yes” and the other says “no”’ (=Ch)). The check category was also coded if a hypothesized synonym was substituted in a part of the sentence which originated from the text (an example for the word ‘inventive’: ‘Smart’ [$=I_1$]…‘You should be…smart’ [=Ch], where the fragment ‘You should be…’ originates from the original text). The categories Accept (Ac) and Reject (Re) pertain to the evaluation of the generated meanings. Acceptance or rejection of an inferred meaning was coded only if the utterance contained an explicit marker, such as ‘Yes, it is [answer]’, ‘It is…’ (example: ‘Yes, improvising it is’; ‘No, it is not…’). The check, accept and reject codes always pertained to the immediately preceding hypotheses and were not indexed. A Residual category (Res) was included to classify utterances that were not related to the actual process of deriving word meaning from context or could not be categorized otherwise. Utterances from the residual category were excluded from the final sequences. Finally, identical codes could follow one another in the coded protocols.

The 360 protocols, which did not contain any information about the students, were coded by two independent coders. One of the coders was informed of the research hypotheses, whereas the other was blind to them. Coefficient kappa $K_n$ (Brennan & Prediger, 1981) and the percentage of agreement were computed for each distinguished activity: inferring ($K_n = 0.78$, percentage agreement = 82%); check (0.59, 67%), accept (0.76, 81%), reject (0.88, 90%), and residual (0.83, 86%). Agreement between observers for the indexation of inferred word meanings proved adequate (0.83, 86%). Disagreements were resolved through discussion.

### *4.5. Scoring of answers*

As deriving word meaning from context is an incremental learning process, partial word knowledge was credited in the evaluation of the students’ definitions. The quality of the concluding answers (CA) was evaluated on three aspects. First, the percentage of correct attributes (PCA) was scored, based on the number of semantic attributes specified in the dictionary definition and the number of attributes present in the students’ definition. For example, the word ‘lanky’ was defined as ‘long and thin’ in the dictionary. If a student defined ‘lanky’ as ‘thin’, only one out of the two relevant attributes for this target word was included, hence a 50% score. An answer was considered fully correct (100% score) if the answer contained all correct attributes or if a synonym for the target word was provided.

Second, the presence of a false attribute (FA) was scored. A false attribute was defined as an attribute that was not listed in the specification of correct attributes. False attributes proved easy to categorize, generally speaking, because they were not at all related to the target word. This aspect was scored dichotomously (i.e. the answer did or did not include a false attribute).

Finally, the aspect contextualization (CTXT) indicated that a definition was not abstracted from the original context. For example, the answer 'thin fingers, so that you can reach into any crack or so' is a contextualized definition of 'lanky' if the words 'fingers' and 'crack' are mentioned in the original context and return in the given definition. The answers 'limber and thin' are considered a decontextualized definition because a detached definition is formulated that does not include words from the original context. Contextualization of a definition is dichotomously scored (i.e. contextualized or not contextualized).

The reliability of the coding of definitions was determined by computing $r_{\mathrm{I}}$ (design 2) (Shrout & Fleiss, 1979) for the interval category PCA and $K_n$ (Brennan & Prediger, 1981) for the nominal categories FA and CNTX. Agreement was satisfactory for correct attributes ($r_{\mathrm{I}} = 0.93$; agreement = 91%), inclusion of false attributes ($K_n = 0.80$, 90%) and contextualization ($K_n = 0.79$, 90%).

### *4.6. Analysis*

The think-aloud protocols were analysed on three dimensions, namely separate activities, transitions and sequences. First, it was analysed whether the protocols showed activities that were not predicted and, conversely, whether predicted processes were not shown (van Someren, Barnard, & Sandberg, 1994). Second, categorical sequence analysis (Bakeman & Gottman, 1997; Gottman & Roy, 1990) was used to identify the way in which the process of deriving word meaning from context was patterned. At the core of sequential analysis is the examination of whether particular transitions of behaviour categories occur more or less frequently than would be expected to occur by chance (Sackett, 1978). In the present study, dependency was formulated in terms of first-order probabilities and correspondingly adjusted residual $z$-scores (Bakeman & Quera, 1995) were computed for transitions from one code to the immediately following code. Consecutive codes could repeat in the transition matrix.

The analysis was focussed not only on single transitions from one step to the next, but also on the full sequences. Similarity between sequences was determined using the agreement measure $r_{\Upsilon}$ (Dijkstra & Taris, 1995). This measure, which is conceptually similar to a bivariate correlation measure, is based on the number of alterations that are necessary to turn a particular sequence into a sequence with which it is compared. The measure is standardized by relating the number of alterations to the number of moves needed to turn the sequence into its reverse order; the values range from 0 (no elements in common) to 1 (identical sequences). For example, the agreement between the sequences ABCA and ACBA – where the codes are identical but the order differs – is 0.86, and a comparison between the ABCA sequence and the sequences ADCA (one new code) and ABEB (two non-identical codes) shows a decreasing agreement of 0.75 and 0.50, respectively. The measure, which is sensitive to differences in length, can also indicate the similarity between sequences with different numbers of elements (e.g. $r_\gamma$ is 0.86 for ABC–ABDC). Subsequently, it was analysed whether sequences could be clustered based on their similarity scores. The agreement criterion for a cluster was set at 0.70 to detect relatively homogeneous clusters. A second criterion was that a cluster describes 10% or more of the protocols to avoid a proliferation of small clusters. Finally, the relation between the resulting clusters and answer quality was studied. The sequence and cluster analyses showed the same results for all grades, and the data were therefore pooled to summarize the findings. It is reported how many students from the sample reflect a certain event, transition or cluster to indicate its generality (Bakeman & Gottman, 1997).

Significance tests were performed at $\alpha = 0.05$. The alpha level was set at 0.001 for the lag sequential analysis, because multiple tests were performed on the transition table. $Z$-scores above the cut-off score 3.29 ($p < 0.001$) indicate a transition that occurs relatively frequently, considering the marginal frequencies.

## 5. Results

A total of 1386 events (plus 206 residual events) were identified in the 360 protocols. The number of transitions was 1026. The average sequence comprised 3.85 events (sd = 1.83), including the final answer. Hypotheses predominated in the protocols (61.2%), excluding the concluding answers. Checks (25.5%) and acceptation of an inferred word meaning (12.4%) were also observed, while rejections rarely occurred (0.8%). The number of different inferred meanings for one target word ranged from one to six. A summary is given in Table 1, broken down by grade.

The model seems adequate in describing the kind of activities students employ in deriving word meaning from context. The percentage of utterances that were classified as residual (13%) is not negligible, but these were mostly activities unrelated to the task. For example, students sometimes made remarks about the difficulty of the task, or directed remarks to the experimenter. A response that was

Table 1
Frequency of coded categories by grade

| | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | Ch | Re | Ac | $CA_1$ | $CA_2$ | $CA_3$ | $CA_4$ | $CA_5$ | $CA_6$ | Res |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grand mean | 381 | 167 | 59 | 16 | 4 | 2 | 262 | 9 | 127 | 199 | 83 | 47 | 24 | 6 | 1 | 206 |
| Grade | | | | | | | | | | | | | | | | |
| 2 | 128 | 45 | 14 | 2 | 1 | 0 | 73 | 4 | 38 | 74 | 30 | 12 | 4 | – | – | 48 |
| 4 | 126 | 56 | 16 | 3 | 0 | 0 | 83 | 1 | 32 | 66 | 24 | 20 | 8 | 2 | – | 53 |
| 6 | 127 | 65 | 29 | 11 | 3 | 2 | 106 | 4 | 57 | 59 | 29 | 15 | 12 | 4 | 1 | 105 |

not anticipated in the scoring system was the students' paraphrasing of large fragments of or even the full paragraph in describing what the unknown word meant. This paraphrasing, which was found in only three protocols of two of the students, shows that the context is considered across sentence boundaries. For example, the following fragment was found in one of the protocols:

> 'So *dilemma* means, for example, that you're doing a test and someone copies you and they keep on doing that in every test, and you're fed up with it, and you want to tell the teacher, but you know that the other person will get angry, and so you wonder should I tell the teacher or not. You just don't know what to do about it, that it…'.

All protocols started with an inferred meaning for the unknown word. This outcome, which may partly be occasioned by the think-aloud procedure, fits the process model. The generation of a word meaning was most often followed by a check (167 times) or the generation of a second word meaning (102 times), although acceptance of the meaning was also observed (35 times). In 55 protocols, the sequence consisted of a single word meaning which was repeated as the concluding answer.

Interestingly, the students verified inferred word meanings associated with the check sequences more often with concrete than with abstract words (58 vs. 88%), $F$ $(1, 358) = 10.31$, $p = 0.001$. They also generated more possible meanings with concrete words than with abstract words (average scores of 1.63 and 1.87, respectively), $F$ $(1, 358) = 5.06$, $p = 0.025$. Acceptance of ideas occurred equally often for the two word types, whereas rejection was found more often for concrete words, although the difference could not be tested meaningfully because of their infrequent occurrence. To conclude, students generate more ideas and check them more often when deriving the meaning of concrete words from context.

### 5.1. Lag sequential analysis

As stated, the following transitions are predicted by the process model: inferring a meaning, followed by a check ($I_i \rightarrow Ch$), check and accept ($Ch \rightarrow Ac$), check and reject ($Ch \rightarrow Re$), accept and concluding answer ($Ac \rightarrow CA_i$), reject and new inferred meaning ($Re \rightarrow I_{i+1}$), and, finally, consecutive meanings ($I_i \rightarrow I_{i+1}$).

The transition from a first inferred meaning to a check ($I_1 \rightarrow Ch$) occurred frequently, as expected, and was found in the protocols of all the students. The transitional probability was 0.45 with a corresponding $z$ value of 10.92. The proportion of transitions from subsequent hypotheses to a check was not significant, however. The predicted contingency between a check and acceptance was not frequently observed either ($Ch \rightarrow Ac$: 0.15, $z = 1.43$), and was reflected in the protocols of only 19 students. Apparently, students do not often check their hypotheses and then explicitly confirm or reject them. Rejection of a hypothesis was observed only nine times, which precluded the meaningful testing of significant transitions. Acceptance of an answer was frequently followed by a concluding answer ($Ac \rightarrow CA_1$: 0.58, $z = 10.66$) and was found in the protocols of every student.

As expected, significant transitional probabilities were also observed for the successive formulation of different hypotheses, and the general pattern where one hypothesis is followed by the next proved universal. As a result of inhibition by the frequent $I_1 \rightarrow Ch$ transition, the proportion of $I_1 \rightarrow I_2$ transitions was not significant (0.28, $z = 2.04$), although such a transition was observed in the protocols of every student. The transitions $I_2 \rightarrow I_3$, $I_3 \rightarrow I_4$ and $I_5 \rightarrow I_6$ occurred relatively frequently (0.35, $z = 11.60$; 0.39, $z = 14.34$; and 0.25, $z = 9.17$, respectively) and only the $I_4 \rightarrow I_5$ transition was not significant. Of all transitions, 58% can be explained directly by the process model. The transition from a check to a second hypothesis, which was not predicted, was also statistically significant ($Ch \rightarrow I_2$: 0.42, $z = 7.58$). No other statistically significant transitions were observed.

The differences in predicted transitions between concrete and abstract words only approached statistical significance ($p = 0.06$). This trend is mainly the result of the fact that the $I \rightarrow Ch$ transition was observed more often for sequences with concrete words than with abstract words (83% and 57%, respectively).

To summarize, the predicted transition from an inferred word meaning to a check occurred frequently, particularly when deriving the meaning of concrete words. Second, the students often formulated a final answer after confirming their inference, as expected. They also sometimes ‘brainstormed’ (i.e. inferred several meanings in a row), as evidenced by the significant transitions for successive guesses.

### *5.2. Cluster analysis*

As the process model did not seem to give an exhaustive account of the data, a cluster analysis was carried out in order to describe the process of deriving word meaning from context in further detail. The majority of the protocols (81%) could be clustered into four relatively homogeneous groups. The first cluster consisted of short $I_1 \rightarrow CA_1$ sequences, which occurred 55 times in the protocols. This cluster describes the sequences where a student formulated a word meaning that was repeated as the concluding answer, as in the following protocol: ‘Dilemma is sort of a problem’ [long pause]...[Experimenter:] ‘What definition do you think the dictionary would give for this word?’...[Student:] ‘Problems’. This type of sequence can therefore best be characterized as a hypothesis-only sequence. All sequences were identical and therefore the mean agreement for this cluster is 1.

A second cluster of 50 sequences is best described as an $I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow Ac \rightarrow CA_1$ sequence (mean agreement: 0.76). This cluster is characterized by the formulation of several word meanings, and has therefore been labelled a brainstorm sequence. Inspection of the protocols revealed that students develop several hypotheses in a row, which do not necessarily build upon each other. A case in point is the following protocol for ‘facade’: ‘A wall or something like that...or a fence...It can also be a cycle rack...Yes, I think it’s a cycle rack’ [pause]...[Experimenter:] ‘What definition do you think the dictionary would give for this word?’...[Student:] ‘I think, wall’. Another student jumped from ‘shed’ to ‘bike’ and then to ‘rain’ for the same target word, but finally reverted to ‘shed’. This observation is in line with Harmon (1998), who notes that students ‘toyed around with various meanings as they tried to make sense of targeted words' (p. 586). Although older students generated more word meanings, as expected, the number of brainstorm sequences was nearly equal for the three grades, with frequencies of 16, 20 and 14 for grades 2, 4 and 6, respectively. Thus, although on average younger students generate fewer hypotheses, they are certainly not bound to just one hypothesis.

The most frequent cluster (150 sequences) was observed in the protocols of each student and is best described as an $I_1 \rightarrow Ch \rightarrow I_2 \rightarrow CA_1$ sequence (mean agreement: 0.82). An important feature of this cluster is the check of the first hypothesis, which is followed by another hypothesis. Sequences from this cluster can therefore be summarized as check sequences. An example for 'inventive' is the generation of the hypothesis 'on the ball', which was subsequently checked by substituting it for 'inventive' in the original sentence ('You should be…on the ball').

Finally, a small cluster was made up of 35 sequences that are identical or similar to an $I_1 \rightarrow I_2 \rightarrow Ch \rightarrow CA_3$ sequence (mean agreement: 0.81). The sequences of this smaller cluster were found in the protocols of 19 students. In these sequences, students infer several meanings, perform a check, and finally prefer a new meaning. Sequences from this cluster, which are typically finished with a new meaning, can therefore be summarized as check-and-switch sequences. The protocols show that the definitive answer was sometimes a related but semantically different answer. However, many protocols showed a clear shift to an unrelated concept.

No relation was found between sequence type and grade, $\chi^2$ (6) = 3.78, $p = 0.71$. Nor was any relation found between sequence type and word type, $\chi^2$ (3) = 4.90, $p = 0.18$. Abstract and concrete words were approximately equally represented in each cluster, allowing a test of the relationship between definition quality and word and cluster type.

The four sequence types appeared in the protocols of the majority of the students (hypothesis-only: 73%; brainstorm: 67%; check: 100%; check-and-switch: 63% of the students). Furthermore, these sequences made up the greatest part of sequences (67–100%) for most students. Two students showed a deviant pattern, however, with many sequences that could not be clustered (50 and 83%, respectively). Further inspection showed that these sequences were similar to the non-clustered sequences of other students; sequences that could not be clustered were also observed for the majority of students (87%). The sequences were relatively long as a result of the combination of check, switch or brainstorm elements, and, hence, they were not included in one of the clusters. It should also be noted that agreement between sequences often decreases rapidly as they get longer with the cluster technique used in this study, leaving a miscellaneous category. These findings provide some evidence for the generalizability of the distinguished sequence types across students, although they may not be universal.

### *5.3. Relationship between students' activities and answer quality*

How are the four clusters related to the quality of the definitions that students formulate in deriving word meaning from context? And is this relationship similar for abstract and concrete words? An ANOVA was performed for the analysis of the percentage of correct attributes with sequence cluster and word type (abstract or concrete) as factors. The word-type factor approached significance ($p = 0.08$), whereas a significant difference between the clusters was observed, $F(3, 282) = 3.11$, $p = 0.027$. The interaction effect of cluster and word type was not significant ($p = 0.27$), indicating that performance was similar for abstract and concrete words within each cluster. A Tukey post hoc test indicated that the difference between clusters should be attributed to the higher scores from the hypothesis-only cluster compared to the brainstorm and the check cluster, although the difference now only approaches statistical significance with $p = 0.063$ and 0.058, respectively. The check-and-switch cluster had a middle position (see Table 2).

The inclusion of false attributes was significantly related to both cluster type, $\chi^2$ (3) = 11.08, $p = 0.011$, and word type, $\chi^2$ (1) = 6.09, $p = 0.014$. False attributes were more often included in abstract words (69%) than in concrete words (54%). A false attribute was included in half of the answers from the hypothesis-only and brainstorm clusters, whereas this increased to 65 percent for the check cluster and to 80 percent for the check-and-switch cluster. A more differentiated result was found when cluster results were broken down into sequences for abstract and words. The abstractness of a word proved a significant predictor for the check and the brainstorm sequence ($p = 0.005$ and 0.048, respectively), but not for the other two sequences. The relationship between the check sequence and false attributes is strong for abstract words (76% of the cases) and weaker for concrete words (54%). Interestingly, a sharp contrast was observed for the brainstorm sequence, which showed many false attributes for abstract (64%) but not for concrete words (36%), suggesting that brainstorming may be a more effective strategy with concrete words. It should be noted, though, that the sample sizes in this analysis were fairly small.

Finally, the decontextualization of definitions was not related to word type ($p = 0.75$), whereas significant differences were found between the distinguished sequence clusters, $\chi^2$ (3) = 26.42, $p = 0.000$. No significant differences were observed between abstract and concrete words within clusters. The definitions from the hypothesis-only and the brainstorm cluster were more often decontextualized than were those from the check and check-and-switch cluster.

In sum, the hypothesis-only sequence was associated with answers with a relatively high percentage of correct attributes and relatively few false attributes. In addition, most answers from this cluster were decontextualized. The brainstorm sequence ranked after the hypothesis-only sequence. Although this sequence was associated

Table 2
Means for percentage of correct attributes (PCA), and proportion of answers with false attributes (FA), and contextualization (CTXT) by cluster; standard deviation is given between parentheses

| Cluster | *N* | Measure | | |
|---|---|---|---|---|
| | | PCA | FA | CTXT |
| Hypothesis-only | 55 | 51.7 (43) | 0.51 | 0.22 |
| Brainstorm | 50 | 32.9 (37) | 0.50 | 0.12 |
| Check | 150 | 36.3 (36) | 0.65 | 0.47 |
| Check-and-switch | 35 | 41.4 (44) | 0.80 | 0.43 |

with a relatively low percentage of correct attributes, the definitions did not often include false attributes and were predominantly decontextualized as well; abstract and concrete words showed different results, however. The definitions resulting from the check and check-and-switch sequences were typified by the inclusion of false attributes in highly contextualized definitions. Students check their hypothesis by fitting it in the context, but apparently then experience difficulties in becoming detached from it. This conclusion is also supported if all sequences are analysed. In this analysis, the presence of a check in a sequence was not associated with the percentage of correct attributes in the concluding answer, $F(1, 358) = 0.22$, $p = 0.64$, but was related to the inclusion of false attributes, $\chi^2$ (1) = 5.44, $p = 0.02$, and contextualization, $\chi^2$ (1) = 32.9, $p < 0.001$. Finally, cluster type showed stronger relationships with definition quality than with word type, although the quality of students' definitions was significantly lower for abstract words, as indicated by the larger proportion of false attributes.

## 6. Conclusions and discussion

Primary-school students employ several activities and encounter various problems in deriving the meaning of conceptually challenging words from a written context. They infer one or more meanings, check them and then evaluate their perceived quality to arrive at a final answer. Students from all grades substituted their inference for the unknown word in the sentence to check the grammatical and semantic fit of their tentative answers. Typical problems are finding the meaning and formulating ideas in the inference phase, using the grammatical and semantic information in the checking phase, and selecting a concluding answer in the evaluation phase if several meanings have been generated.

The activities that showed up in the protocols are not performed in a fixed series. Students do not pass through an invariable sequence of generating, checking and evaluating for each meaning: they often take shortcuts, thereby omitting any orientation and evaluation activity. A possible explanation for this is that students know a number of tactics, but apply them flexibly. Only if a first tentative guess is considered an unlikely candidate or one that can be improved upon may a student decide to undertake further action, namely to check the hypothesis and/or generate a new one as a 'repair tactic'. This also suggests that the process of deriving word meaning from context can be characterized as product-oriented, while evaluation activities seem to be employed on an ad hoc basis.

### *6.1. Research implications*

The process model used in this study proved valid in describing separate activities. In addition, fit was achieved on event and transition level, although not all the sequential implications of the model are fully supported by the data. The descriptive value of the model would be strengthened by making post-inference evaluation activities optional. This study suggests that students do not perform evaluation activities automatically but apply them as 'repair strategies', that is, when they are in doubt or need a better answer.

Furthermore, the role of contextual analysis could be given more focus in future think-aloud studies on deriving word meaning from context. Adding the paraphrasing of small or large fragments of the text as a coding category would improve the descriptive quality. However, other measures may also be fruitful to reinforce the role of context use in think-aloud studies, which may be underexposed due to its automatization. Students could, for example, be instructed to underline useful parts of the context during the task or be invited to motivate their answers retrospectively.

In interpreting the results of this study, three factors should be kept in mind. First, the commonly used design where identical materials are presented to students of different ability and grade does not exclude the possibility that the experimental task is more difficult for the youngest students, which may affect their cognitive processing. Second, the data were pooled across a relatively small sample of students and target words for inferential tests, which limits the generalization of the findings (Bakeman & Gottman, 1997). The descriptive statistics showed that the distinguished events, transitions from one step to the next, and the distinguished clusters were generally found in the majority of the protocols, but further research is necessary to establish the generality of the results of this study to other students, and to other words and other contexts. The product measures used in this study are the final concern. Vocabulary learning is considered a slow, incremental process (Nagy & Scott, 2000). Students can thus not be expected to give a dictionary-like definition of an unknown word after just one encounter with it; providing multiple contexts do not solve this problem, as previous studies have shown. Tests that measure the skill of deriving word meaning from context typically show modest students' scores, as in this study, although partial word knowledge is credited. In addition, finding the meaning of an unknown word is secondary to constructing meaning for the text in normal reading. Seen from this perspective, finding the single most appropriate meaning for an unknown word and subsequently carrying on with reading is an efficient strategy.

Finally, many authors have pointed out the significance of think-aloud data for studying the processes in which readers are engaged (Afflerbach, 2000; Kucan & Beck, 1997; Pressley & Afflerbach, 1995). This line of research would benefit from the application of sequential analysis, because it is suited to describe the sequential nature of cognitive processes. Sequential analysis was not found by Pressley and Afflerbach in their review of reading-process research, nor is it common in other cognitive domains (but see Neuman, Leibowitz, & Schwarz, 2000), although it has been successfully applied in the analysis of processes in other scientific disciplines such as sociology (event history analysis), psychology (developmental studies), biology (animal behaviour), linguistics (discourse analysis) and educational science (observational studies on classroom behaviour and human-computer interaction).

### 6.2. Instructional implications

The results of this study lead to a few cautious suggestions for effective contextual analysis instruction, taking into account its limitations. Many authors have emphasized the need to check an inference in deriving word meaning from context. Fischer (1994: 569), for example, argues that 'learners could avoid overgeneralizing unfamiliar words if they initially tied their understanding of a novel word closely to the context in which they first encountered it'. However, in this study the checking of hypotheses did not lead to superior definitions. It appears that when students rely too heavily on context, they undergeneralize the meaning of unknown words by giving highly contextualized definitions and including irrelevant features. Thus, the substitution tactic certainly seems to be of value, but it has a drawback. This suggests that students should learn not only how to use context but also how to become detached from it.

Students' natural process of deriving word meaning seems flexible. This suggests that teaching complex strategies may not be the best approach and may explain why relatively simple heuristics have proved successful, whereas elaborate strategies involving four to six steps were less effective (see Fukkink, 2002; Fukkink & de Glopper, 1998). The feasibility of an instructed approach is particularly important if we want to promote transfer from a strategy for deliberately deriving word meaning from context to incidental word learning during normal reading. However, weaknesses in students' tactics should be acknowledged. It therefore seems particularly helpful to supplement the creative generation of hypotheses in deriving word meaning from context with an increased understanding of orientation and evaluation activities.

## Acknowledgements

The author wishes to thank three anonymous reviewers for their helpful comments on earlier versions of this manuscript.

## Appendix A

Target words (translated from Dutch)
Abstract words: dilemma; suspicion; to boast; to feign; inventive; resolute.
Concrete words: facade; tirade; to subside; to shatter; lanky; surreptitious.

## Appendix B

Two examples of texts (translated from Dutch)

*Copying*

Loes had copied during the test. Fieke knew this for sure. Loes had often copied before.

Fieke told her parents all about it. ‘I don’t want Loes to copy,’ she said. ‘What should I do? Should I tell the teacher? But Loes would get angry.’

Her dad nodded.

‘Or should I ask Loes not to copy?’ Fieke asked.

Fieke had done this already, however. Loes had become very angry. ‘I don’t copy!’ she had shouted.

‘Well,’ her dad said, ‘that is a *dilemma*.’

He mum did not know what Loes should do: ‘I don’t know what else you can do either. You know best. Think about it carefully.’

*Bert and Kees play draughts*

Bert and Kees wanted to play draughts. They looked for the draughtboard. It was in the cupboard. But where were the draughts? They looked everywhere. The draughts were gone.

Dad could not find them either. But he had a solution. ‘You should be *inventive*,’ he said.

He picked up a mandarin and peeled it into many pieces.

‘Watch,’ dad said. ‘These peels are for Bert: they’re orange.’ He laid them on the draughtboard, then turned over the other peels. ‘These are for Kees: they’re white.’ Bert played using the orange peels. And Kees played using the white ones. Now they could play draughts.

‘Don’t turn over any pieces behind my back!’ dad called. Kees and Bert laughed.

## References

Afflerbach, P. (2000). Verbal reports and protocol analysis. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research*, *Vol. III* (pp. 153–162). Mahwah, NJ: Lawrence Erlbaum.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: an introduction to sequential analysis* (2nd ed.) Cambridge: Cambridge University Press.

Bakeman, R., & Quera, V. (1995). Log-linear approaches to lag-sequential analysis when consecutive codes may and cannot repeat. *Psychological Bulletin*, *118*(2), 272–284.

Baumann, F., Edwards, E. C., Font, G., Tereshinski, C. A., Kame’enui, E. J., & Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly*, *37*(2), 150–176.

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., McKeown, M., & McCaslin, E. (1983). All contexts are not created equal. *Elementary School Journal*, *83*, 177–181.

Blachowicz, C. L. Z., & Fisher, P. (2000). Vocabulary instruction. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research*, *Vol. III* (pp. 503–523). Mahwah, NJ: Lawrence Erlbaum.

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).

Breetvelt, I., van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality – when and how? *Cognition & Instruction*, *12*(2), 103–123.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, *41*, 687–699.

*Celex*. Nijmegen: Max Planck Institüt für Psycholinguistik, Centre for Lexical Information.

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

van Daalen-Kapteijns, M., Elshout-Mohr, M., & de Glopper, K. (2001). Deriving the meaning of unknown words from multiple contexts. *Language Learning*, *51*(1), 145–181.

Dijkstra, W., & Taris, T. (1995). Measuring the agreement between sequences. *Sociological Methods & Research*, *24*(2), 214–231.

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: Vol. 51-08A (pp. 2691).

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: verbal reports as data*. Cambridge: MIT Press.

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

Fukkink, R. G. (2002). Effects of instruction on deriving word meaning from context and incidental word learning. *L1: Educational Studies in Language and Literature*, *2*(1), 37–57.

Fukkink, R. G., Blok, H., & de Glopper, K. (2001). Deriving word meaning from context: a multicomponential skill. *Language Learning*, *51*(3), 477–496.

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), 450–469.

Goerss, B., Beck, I. L., & McKeown, M. G. (1999). Increasing remedial students' ability to derive word meaning from context. *Reading Psychology*, *20*(2), 151–175.

Gottman, J. M., & Roy, A. K. (1990). *Sequential analysis: a guide for behavioral researchers*. Cambridge: Cambridge University Press.

Harmon, J. M. (1998). Constructing word meanings: strategies and perceptions of four middle school learners. *Journal of Literacy Research*, *30*(4), 561–599.

Harmon, J. M. (1999). Initial encounters with unfamiliar words in independent reading. *Research in the Teaching of English*, *33*, 304–338.

Jenkins, J. R., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: the teaching of individual word meanings and practice in deriving word meaning from context. *Reading Research Quarterly*, *24*(2), 215–235.

Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: inquiry, instruction, and social interaction. *Review of Educational Research*, *67*(3), 271–299.

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), 119–138.

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.

Nagy, W., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*, 304–330.

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

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

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 269–284). Mahwah, NJ: Lawrence Erlbaum.

Neuman, Y., Leibowitz, L., & Schwarz, B. (2000). Patterns of verbal mediation: a sequential analysis of self-explanation. *Journal of Experimental Education*, *68*(3), 197–213.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: the nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum.

Sackett, G. P. (1978). Measurement in observational research. In G. P. Sackett (Ed.), *Data collection and analysis methods. Observing behavior: Vol. 2*. Baltimore: University Park Press.

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

Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior*, *22*(1), 71–97.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: a practical guide to modelling cognitive processes*. London: Academic Press.

Staphorsius, G., & Verhelst, N. D. (1997). Indexering van de leestechniek. [The development of a domain referenced index of the decoding load of texts]. *Pedagogische Studiën*, *74*, 154–164.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: a meta-analysis. *Review of Educational Research*, *69*(3), 261–285.

Verburg, M., & Huigen, M. (1994). *Van Dale Juniorwoordenboek Nederlands*. [*Van Dale Junior Dictionary Dutch*]. Utrecht: Van Dale Lexicografie.

Waern, Y. (1982). How do you fill in this xxx? On some interpretation process. In A. Flammer, & W. Kintsch (Eds.), *Discourse processing* (pp. 152–165). Amsterdam: North-Holland Publishing Company.

Waern, Y. (1988). Thoughts on text in context: applying the think-aloud method to text processing. *Text*, *8*(4), 327–350.

Werner, H., & Kaplan, E. (1952). The acquisition of word meanings: a developmental study. *Monographs of the Society for Research in Child Development: Vol. 15*, ser. no. 51, no. 1.