Applied Psycholinguistics **22** (2001), 1–33
Printed in the United States of America

Effects of a format-based second language teaching method in kindergarten

NOËLLE UILENBURG and FRANS X. PLOOIJ
Paedological Institute of the City of Amsterdam

KEES DE GLOPPER and RESI DAMHUIS
University of Amsterdam

ADDRESS FOR CORRESPONDENCE
Frans X. Plooij, Paedological Institute of the City of Amsterdam, Ijsbaanpad 4, 1076 CV Amsterdam, The Netherlands. E-mail: fplooij@kiddygroup.com

ABSTRACT
This article concerns second language teaching with a format-based method. The following article in this issue concerns foreign language teaching using the same approach. In the present study, the differences between a format-based teaching approach and a standard approach used as treatments in a quasi-experimental, nonequivalent control group design are described in detail. The hypothesis tested is whether the effects of a format-based teaching method and a standard second language teaching method differ after two years for children at the age of 6. Results of MANOVA analysis for repeated measures show that the groups are quite comparable on the dependent variables, with some small differences in development between the two groups. Univariate results reveal that these differences favor children in the experimental group for sound articulation and favor children in the control group for sound discrimination. The results of the present article and those of the following article are discussed in light of broader issues concerning the differences between second language and foreign language teaching.

The need to teach immigrant children the language of the country as early as possible and the mandate to begin foreign language teaching earlier than before make it imperative to find effective teaching strategies for kindergarten children. The use of drama and music are highly recommended in language teaching, especially for this age group. Taeschner (1991) was able to wed them into a theoretical base, developing an innovative, format-based approach which attempts to bring natural language acquisition processes into the classroom.

The research results reported in the present article and in the following article in this issue were part of the European Project for Teaching a Second or Foreign Language to Preschool and Schoolaged Children, which developed the teaching materials, guided the process of the implementation, and evaluated the effects

© 2001 Cambridge University Press 0142-7164/01 $9.50

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

of the innovative approach in four European countries: England (University of Birmingham), Italy (Universita di Roma "La Sapienza," Rome), Spain (Basque Country: Universidad del Pais Vasco, San Sebastian), and the Netherlands (Paedological Institute of the City of Amsterdam and the University of Amsterdam). For two years, the method was used for second language teaching in England and the Netherlands and for foreign language (English) teaching in Spain and Italy. The project was subsidized by the European Community's LINGUA program. The present article and the article that follows in this issue report the results from the Netherlands and the Basque Country, respectively.

The main objective of the project was to find ways of improving the teaching and learning of languages in Europe. This was achieved through the implementation of, and experimentation with, an innovative, format-based teaching approach. The same scripts, picture books, and songs on audiocassette were used. The supplementary material reflected the national flavor of each country but was still essentially the same. In order to make the same teaching materials available to every country and for any kind of teaching setting, the notion of "supranationality" served as the criterion for the characters and the adventures that were incorporated in the teaching materials. Cultural biases were erased; cultural values were avoided or at least not violated (e.g., certain animals, such as pigs or dogs, that were known to be offensive in certain cultures were excluded). Moreover, any connotation of sexism, suppression, or inequality between characters was deleted from the materials.

In addition, special training in drama techniques and communication strategies was given to the teachers from the different countries. The success of any method depends to a great extent on how it is carried out in the classroom. The conditions of the schools and classes and the linguistic and pedagogic abilities of the teachers are vital factors. In order to provide the best and most homogeneous conditions to carry out the project, initial training in psychological and pedagogic principles, drama techniques, classroom management, and teaching techniques was provided; guidance was also given on how to reflect on classroom practice through a process called action research.

The research consisted of two parts: a product evaluation and a process evaluation. A pretest/posttest nonequivalent control group design was used for the product evaluation. At the beginning and at end of the project years, data on the linguistic performance of the children were collected with discrete language assessment instruments and also in more interactive test situations. For the process evaluation, a mixed time series design was used. At regularly spaced intervals over two years, a series of data was collected on (a) linguistic progression, (b) social interaction, which focused alternately on the teacher and the children, and (c) action research, which explored the teachers' experiences during the project.

All participating research teams strived to make the data collection and data analysis as comparable as possible. Conformity was reached to a high extent but, due to all sorts of unforeseen practical obstacles, was not perfect. Obstacles included malfunctioning equipment and refusal by individual participants to be studied. The two articles presented in this issue detail the local specifics of research design, data collection, and instrumentation.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

THE DUTCH PROJECT

Young immigrant children leave kindergarten with a low Dutch language proficiency as compared to native Dutch-speaking peers (Boogaard, Damhuis, de Glopper, & van den Bergh, 1990). School failure and dropout often occur as a consequence of this. To ensure that these children benefit from their primary school education, it is important to start as early as possible (i.e., in kindergarten) and to use a good second language teaching approach. However, it is difficult to choose the right approach for kindergarten[1] children, because their young age requires an approach that is not easily constructed into a method.

Research on first and second language acquisition has tried to identify which aspects of the language environment are beneficial to the process of language acquisition. Participation in interaction is one such aspect (Damhuis, 1995; Ellis, 1984; Ochs & Schieffelin, 1995).

Developers of second language teaching methods have tried to integrate interaction in meaningful contexts into the work forms. But in developing second language teaching methods, the problem of what is most important – communicating versus learning the structure of a language – has not yet been solved. Appel and Vermeer (1994) referred to this problem as the conflict between code and communication.

The methods for teaching Dutch as a second language, developed in the last decades, constitute systematically ordered programs and are referred to as code-oriented second language methods. The code-oriented approach evolved from the point of view that second language education must be structured and systematic in order to achieve the best results. Here, the conflict between code and communication is quite clear. The explicit goal is the code – usually a target vocabulary to be learned. Vocabulary is seen as an important prerequisite for language proficiency, and on this level a large disadvantage has been found with immigrant children (Appel, 1984; Verhoeven & Vermeer, 1992).

The way in which the goals are worked out, however, shows that it is very difficult to introduce real interaction into teaching methods. Kindergarten education is usually development-oriented. It does not follow strict methods or programs but rather stimulates attempts to a wide range of developmental aspects, which are adjusted to each child as closely as possible. Although teaching methods try to fit in with the traditional kindergarten approach and there is some freedom in ordering certain parts of the method, many kindergarten teachers find the code-oriented approach incompatible with the goals of kindergarten and judge these methods as being too schoolish.

The format-based approach developed by Taeschner (1991) makes a real effort to bring natural language acquisition processes into the classroom. Getting children to communicate in their second language is the main aim. This innovative approach to second language education is more in line with the developmental orientation of kindergarten. Its basic principles are derived from natural first language acquisition, and its teaching methods are communication-oriented and playful. A format-based approach provides children with optimal conditions in which to acquire a second language and abandons the idea that one explicitly has to teach young children what has to be learned.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

In this study, the hypothesis tested is whether the effects on language proficiency of a format-based second language teaching method and a standard method differ after two years of teaching, using a quasi-experimental, nonequivalent control group design. The similarities and differences of the format-based method and the code-oriented method are described next.

FORMAT-BASED VERSUS CODE-ORIENTED APPROACH

Rationale of the format-based method

In the Netherlands, the format-based method is known as the Dinocroc method, after the imaginary species (a cross between a dinosaur and a crocodile) that provides the main characters in the material.[2] It is a psycholinguistic approach to second language teaching. The principles underlying the method have been derived from processes in first language acquisition.

The important principle of the format-based approach is the presumption that children need to develop communicative intentions in their second language in the same way as they do in their first language (Taeschner, 1991). Communicative intentions are a prerequisite to language learning, whether it be a first or second language. Bruner (1975, 1990) described the development of communicative intentions in the first language of babies. Communicative intentions arise in daily routines of shared experiences between an adult and a child. Bruner called these routines "formats." In the development of intentions in the first language, formats are frequently recurring day-to-day activities, such as feeding, diaper changing, going to bed. Here, the format principle was adjusted for use with school-aged children. Simple, identical repetition of daily routine formats would become boring. Therefore, narrative formats were created in which a chain of events would lead to a final event or climax. The concept of a narrative format refers to a whole storyline and not to the more static format concept, as in Bruner's research. The chains of events are more or less predictable. This increases the involvement of the children. The predictability of the formats is ensured by choosing topics from the children's world of interest (i.e., things they know, things that have happened to them, or things that arouse their emotions). Fantasy play is the means through which the children and their teacher experience the formats.

A prerequisite to the development of communicative intentions in the second language is an affective interaction between the child and the teacher, which is achieved through a good communicative relationship. Intersubjectivity and turn taking in a conversation are important features of a good communicative relationship. These features are incorporated into the way the narrative formats are constructed.

Another aspect in developing a good communicative relationship is that the speakers set and adjust their language production according to what they believe their interlocutors are able to understand. In second language teaching, this means that the teachers must not only adjust their language to the children, but also pretend not to understand the mother tongue. If the teacher did respond to the child's first language, the need for the child to speak in the second language

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

with the teacher would not exist. Starting to communicate in the second language is the main aim of the format-based approach.

In writing the formats, the use of target vocabulary lists and prearranged syntactic structures is deliberately abandoned. The topics of the format and the accompanying natural communication determine the vocabulary and syntactic structures.

The format-based method in the classroom

A total of 20 narrative formats were written to resemble theater scripts. The scripts provided dialogues, stage directions, organization of the classroom, sounds, gestures, and facial expressions to support the language; they also included a description of the characters.

To illustrate a script and its use, we describe one format in more detail. In format 6, Hocus (the male dinocroc) goes out to do some shopping. He explicitly forbids Lotus (the female dinocroc) to go outside because it is raining. When it stops raining, Lotus decides to go out anyway. She plays with her ball, and soon she is covered with mud from head to toe. She goes back into the house in this state, and as a result the whole house gets dirty. Hocus returns, and Lotus tries to hide herself. Of course Hocus discovers her and sees why the house is dirty, and Lotus has to clean up the dirt by herself.

The storyline is predictable and undoubtedly familiar to children. The acting out of the format script is different from normal theater because each child plays all the characters involved in the format, constantly shifting from one character to another, even within a dialogue. In this way, the children experience both sides of a dialogue, and the principle of turn taking is explicitly presented in the formats. Furthermore, the children are active and physically involved all the time, which suits their preference for psychomotor activities and reduces the risk of becoming bored while waiting for a turn.

The use of concrete materials is not encouraged. The acting out itself has to make the concepts clear without the use of visual references. Sometimes tables and chairs are used to mark a certain space. At the beginning of a lesson, the teacher organizes the classroom furniture in such a way as to create an empty floor. The teacher then gives the children some instruction about the organization that is specific to the current story: for instance, "Behind this row of chairs is the house of Hocus and Lotus" or "This table is a bus stop." After the instruction, the acting begins.

For the first two or three times, the teacher sets the example, and the children join in. In the beginning, children who cannot speak Dutch at all can join in the format play without participating verbally. Later, the children and the teacher play the characters simultaneously or the teacher leaves the initiative to the children. In this way, the teacher can adjust his or her role according to the children's knowledge of the format storyline and the language therein.

Each format is acted out by the whole group 12 times (i.e., four times a week for three weeks). The format principle indicates the importance of repetition. The acting out of a format takes 10 to 15 minutes. The 20 scripts cover a teaching period of two years. This means that 20 to 30 hours per year are

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

dedicated to format play. The lexicon covers approximately 700 unique words (nouns, verbs, adverbs, adjectives, prepositions, numerals, and interjections). These are not preselected target words but the vocabulary needed to write the scripts. This means that the vocabulary in the formats is not chosen from the vocabulary lists for kindergarten children, which is the case in the code-oriented approach.

Apart from the formats, there are storybooks, songs, and supplementary materials to work with after acting out the formats. Either the teacher reads aloud from the storybook or the children tell the story with the aid of pictures. The children can work with the storybook individually or in small groups, listening to the mini-musicals or radio plays of the same story.

The mini-musicals are sung by the whole group. All children take a tape of the mini-musicals home with them. There they can listen to the story and experience the adventure again. In this way, their contact with the formats and the Dutch language is extended further.

During the supplementary part of the lesson, all sorts of activities are used to trigger Dutch language production. The children work in small groups, and the teacher gives individual attention to each group. During these activities, there is spontaneous interaction between children and between children and teacher. This gives the children an opportunity to take the initiative in interacting. The four segments of a lesson together amount to 45 minutes.

In addition to the format-based method, all sorts of usual language activities in kindergarten can be used. The formats provide teachers with topics such as contrasting concepts, names, and functions of body parts that can be developed during other (language) activities. However, no precise program is supplied for these activities.

Rationale of the code-oriented method

The code-oriented method used in the control treatment consisted of two separate, though related, methods: "Knoop het in je oren" (KNOOP 'Listen') (van Kleef, 1988) for 4- to 5-year-olds in the first year of kindergarten and "Laat wat van je horen" (LAAT 'Speak') (van Kleef & Boers, 1990) for 5- to 6-year-olds in the second year of kindergarten. The methods were developed to supplement normal language activities in kindergarten. Principles of receptive approaches such as the "total physical response" (Asher, 1969), the "comprehension approach" (Winitz, 1981), and the "natural approach" (Krashen & Terrell, 1983) inspired the construction of the methods.

In KNOOP the focus is on the extension of receptive vocabulary, based on the language acquisition principle that comprehension precedes production, whereas in LAAT the focus extends to productive vocabulary as well. The material was organized around certain themes, and a lexicon was assembled around each of the themes. The final selection criterion for the lexicon for both KNOOP and LAAT was the presence of the items from the New Target-List of Vocabulary for 6 Year Olds (Kohnstamm, Schaerlaekens, de Vries, Akkerhuis, & Frooninckx, 1981). In this way, a target vocabulary of 270 words (nouns and verbs) for KNOOP and 448 words (nouns, verbs, adverbs, and adjectives) for LAAT

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

was assembled. The relevance of the items in the vocabulary was checked through teacher judgment during a pilot study of the implementation.

In KNOOP, apart from receptive vocabulary, listening proficiency gets a lot of attention. Getting children to repeat lexical items and to offer their meanings in different contexts and through different activities is described in the manual of KNOOP as the main aim. These two receptive domains are seen as the basis of later speech proficiency. In LAAT, apart from productive vocabulary, stimulating listening proficiency and speech proficiency is the main aim, but training in receptive vocabulary is important as well. From the description it is clear that the code (vocabulary) is important in the construction of KNOOP and LAAT. The contents of the lessons are determined by the vocabulary that is selected around certain themes.

The code-oriented method in the classroom

The control groups worked with KNOOP during the first year and with LAAT during the second year.

KNOOP contains 10 themes, which can be divided into three phases. Each phase assumes a mastery of the vocabulary of the previous one. The phases show an increase in difficulty. Each of the themes incorporates comparable materials: stories to be read aloud by the teacher, audiocassettes with stories, word cards, murals, and comics of the stories. Each theme takes about three weeks, when working with the method at least three times a week for approximately 20 minutes. This adds up to 30 hours per year. A boy and a girl from a Moroccan family are the main characters in the method.

The manual gives descriptions of the materials, work forms, and vocabulary lists and suggests the use of objects to support the vocabulary. In addition to the target vocabulary, each theme provides extra vocabulary (mainly nouns and verbs, but also adverbs, prepositions, and adjectives) that can be taught to the children (also selected from the New Target-List Vocabulary for 6 Year Olds). The target vocabulary, the lexicon in the stories, and the extra vocabulary together amount to approximately 500 words.

There are group lessons and individual lessons. In the group lessons, the teacher introduces a new theme, reads the stories, and discusses the mural and word cards with the children, who sit in a circle. In the individual lessons, the children listen to stories on a cassette player. Psychomotor language activities are used in at least one lesson per theme.

LAAT contains 8 themes. As with KNOOP, the themes can be divided into three phases. Again, each phase assumes a mastery of the vocabulary of the previous one, and the activities are intended to stimulate speech proficiency of increasing levels of difficulty. Each theme incorporates comparable materials. The mural, new vocabulary, and the stories (listened to on the cassette player or read aloud by the teacher) are recurring aspects of each theme. Apart from this, each theme contains different activities that focus on the repetition of vocabulary and on the stimulation of listening and speaking proficiency. The duration of a lesson is approximately 20 minutes. The manual suggests three lessons per theme per week over a three-week period. This amounts to 24 hours per

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

year. The main characters are five children with different ethnic backgrounds. The manual gives descriptions of the materials, work forms, and vocabulary lists. As in KNOOP, the manual also offers extra vocabulary. When a teacher decides to work with the extra vocabulary, one theme requires three sessions per week over a four-week period. Each theme contains nine group lessons. In each group lesson, the teacher introduces a new theme, reads the story, and discusses the mural, with the children seated in a circle. Apart from the work forms, there are various activities where the vocabulary is presented again and where speech proficiency is a central point of attention. LAAT contains remedial activities for children with very poor language proficiency. In LAAT, the use of concrete materials is encouraged. Each theme ends with a drama lesson. The teacher reads the story and then the children enact it. This drama lesson is repeated twice. The primary goal of the drama lesson is to stimulate speech proficiency and to enlarge the children's understanding of the language.

The manuals for KNOOP and LAAT offer suggestions for teacher behavior during language activities. A teacher should speak slowly and in short and simple, but grammatically correct, sentences. The child should feel at ease during the lessons; therefore, no child should be forced to speak, and no explicit attention should be paid to errors. Correct utterances should be rewarded with a compliment. Implicit correction of errors is advised, as well as repetition of the child's utterance with extension. The teacher must use topics that are of interest to young children. Gestures and mimicry should support the teacher's utterances. Some attention is paid to the kind of answers certain questions invoke as well as the language functions (prediction, logical reasoning, etc.). The manuals also describe how the teacher can evaluate a child's language proficiency.

Comparison of the two methods

Obviously there are similarities and differences between these approaches. The two major differences between the methods are highly related.

First, there is a difference in the linguistic approach. The KNOOP and LAAT methods are centered around the code of the language, whereas the format-based method is centered around communication. Although both methods subscribe to the relevant natural language acquisition principles, different goals are chosen.

The stories used in both methods derive from the world of interest of young children, but in KNOOP and LAAT these stories are not really experienced by the children; they merely sit and listen to them. From the belief that comprehension precedes production, it follows that children would not have to produce while working with KNOOP. They could simply listen to the second language and pick it up.

A big difference between this silent period and the silent period children go through in natural situations is that in natural situations the children are part of the process (i.e., part of a social system). They experience the activities, are actively involved, and are able to understand the words from the context. This is the case with the format-based approach; in contrast, with the code-oriented approach, the children predominantly listen to the stories and only sometimes experience the activities. Children who do not understand the spoken words of

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

the lessons in the code-oriented approach are not really part of the social process and thus are excluded.

The linguistic principles of KNOOP and LAAT resemble the process of natural language acquisition far less than do those of the format-based method. In the latter, vocabulary and the code of language are acquired in the natural process of interaction. In KNOOP and LAAT the language progression is constructed in the materials, whereas in the format-based method the children construct the language progression within themselves. The teacher provides the conditions in which the children can actively build their second language. In KNOOP and LAAT, the vocabulary and the code of language are explicitly taught.

The second major difference concerns the didactic principles in general. The work forms used in KNOOP and LAAT are more traditional and academic and do not match the didactic principles of kindergarten education, which involve learning through play, whereas the didactic principles of the format-based method fit in very well with those of kindergarten education. In KNOOP and LAAT, psychomotor activities and drama are used only occasionally. In the format-based method, the acting out of material is the central work form, whereby children learn along the way and thus actively construct their knowledge of the second language.

The differences highlighted here raise the question of which method is more beneficial to second language acquisition in kindergarten: KNOOP and LAAT or the format-based method.

METHOD

Design

Using a quasi-experimental, nonequivalent control group design (Cook & Campbell, 1979), we wished to find out whether the effects of the format-based method (Taeschner, 1991) and those of the code-oriented method would differ after two years of teaching.

Experimental schools and groups

Children were selected from eight kindergarten classes in three schools with a high number of immigrant children between the ages of 4 and 6. The criteria for selecting experimental schools were as follows: (a) a total school population of more than 200 children and with more than 50% immigrant children; (b) the existence of an internal counseling system in the school. The internal counselors fulfilled two tasks in the project; they provided regular teacher guidance during the implementation, and they helped out in parts of the data collection. The motivation for the schools to join the project was to find a suitable method for teaching Dutch as a second language in kindergarten classes that contain immigrant children of various origins.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Control schools and groups

We selected ten control classes from four schools in which another method for teaching Dutch as a second language was used equally intensively as in the eight experimental classes. For each of the experimental schools, a control school was chosen. These control schools were located in the same area of the city and had the same population as the experimental schools. The methods used in the control classes were KNOOP and LAAT. Logbooks were used to determine the real intensity with which the method was actually used in each class.

Post hoc, the null-hypothesis was tested statistically to see if there were any significant difference between experimental classes and control classes, as far as a number of covariates were concerned.

Selection of subjects

Five children were selected in each class (for the experimental group and for the control group). Selection criteria were (a) a very low level of Dutch language proficiency, (b) normal intelligence, (c) no hearing problems, (d) one out of five children showed internalizing problem behavior, and (e) one out of five children showed externalizing problem behavior. Criteria (d) and (e) require some explanation. From the description of the didactic approach, it is clear that this method would be influenced by the presence of children with problem behavior. In the case of internalizing problem behavior, where children are inhibited to join in the drama, and in the case of externalizing problem behavior, where children disturb the process, reducing their own language learning opportunities as well as those of their playmates, the effectiveness of the approach can be affected. In order to study the effects of these problem behaviors, children with internalizing or externalizing behavior were deliberately selected for the research group. Although the results of this specific aspect of the research are not reported in the present article, the entire selection procedure is described here for the sake of completeness.

The selection procedures were as follows. Two months after the start of the school year, 18 teachers were asked to select five children, aged 4 to 4½, who spoke or understood very little or no Dutch. One of the five children should show "disruptive" behavior, one should be "extremely withdrawn," and the other three should have no problems whatsoever.[3]

In the control groups, the number of selected children did not always equal five. This was because, in the control classes, the number of children who did not understand any Dutch was not equally divided over the kindergarten classes; in some groups there were more than in others. This factor was thought not to disrupt data collection and analysis and was therefore accepted. By the end of the first year, the total number of children in the control treatment was 37 (vs. 40 in the experimental treatment). Due to migration, the number of children dropped to 29 and 32, respectively, by the end of the second year.

To confirm the low levels of language proficiency in the target children, teachers kept a linguistic episode diary. For each target child, during the first two months of the school year, the teachers tried to write down 15 to 30 utter-

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

ances (literally and including all mistakes), along with the context in which these occurred. Based on the diaries, a group of target children was selected that produced very few or no Dutch utterances.

The teacher judgments of the problematic behaviors were verified using the Teacher Report Form (TRF) of the Child Behavior CheckList (CBCL) (Achenbach, 1988). If there was a discrepancy between the teacher's clinical judgment and the CBCL-TRF, another child was selected until judgment and questionnaire matched. It was not possible, however, to select one child with internalizing problem behavior and one child with externalizing problem behavior in each group. In some classes the number of children with internalizing problem behavior was higher, and in other classes the number of children with externalizing problem behavior was higher.

Finally, we checked whether the selected children were of normal intelligence (for this, a nonverbal intelligence test was used: the K-SSON [Snijders & Snijders-Oomen, 1975]), and that they had no hearing problems. In both treatments the average IQ was 110, with a range from 80 to 135. It was accepted that in each treatment two children had quite a low IQ (80–85); apart from the fact that testing young children is always difficult, the test that was used was not specifically constructed to test immigrant children. No children with hearing problems were detected.

Pretest, midtest, and posttest measures

To measure Dutch language proficiency, discrete test data were collected as well as data on spontaneous language production.

For the discrete tests, the Taaltoets Allochtone Kinderen [Language Test for Immigrant Children] (TAK; Van de Guchte, Verhoeven, & Vermeer, 1986) was used. This assessment instrument contains tasks that cover a wide range of linguistic domains, including passive and active vocabulary, morphology, sound discrimination, sound articulation, sentence understanding, sentence building, and comprehension of texts. At the start of testing, this was the only test available for the language assessment of immigrant children. The problem was, this test was only valid for children 5 years and older. At the time we began testing, our subjects had not yet reached this age. Therefore, alternative solutions had to be found for the pretest. For the vocabulary domain, a method-bound receptive test and a productive vocabulary test, which were made method-independent in other research, were used.[4]

For the sentence structure domain, no alternative was available. Therefore, three TAK subtests – Sentence Understanding 2, Sentence Understanding 3, and Sentence Building – were rank-ordered according to degree of difficulty. This adaptation was based on the research of Boogaard et al. (1990) with 6-year-olds. Together with a break-off criterion, the rank-ordered tests were suitable for use with children younger than the minimum age for which the TAK was designed. The other linguistic domains were not tested at the pretest because no alternatives were available.

At the midtest and posttest, the TAK was used completely (with the exception of the vocabulary and sentence structure tests, which were used for the pretest).

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

This was done to permit a comparison over two years. See Appendix 1 for a detailed description of the subtests and how they were used to construct the dependent variables.

This procedure resulted in two sets of dependent variables. One was a set of four dependent variables, which was measured at all three points of testing. This set consisted of receptive vocabulary, productive vocabulary, sentence understanding, and sentence building (Set I). A second set, measured only at the midtest and posttest, consisted of sound discrimination, sound articulation, morphology, and comprehension of texts (Set II).

Spontaneous speech data were collected in a "chat," which the first author had with each selected child. This chat was about pictures of the child's social and physical environment. Pictures of the class, the teachers, and favorite toys were used because certain cultures of our immigrant children prohibited taking pictures of themselves. These pictures were made on the spot using a Polaroid camera. The speech production of the children was audio recorded. Analyses were performed by using the CHILDES program (MacWhinney, 1991).

The language proficiency pretest was carried out at the beginning of the first school year. The midtest was carried out at the beginning of the second school year. The posttest was carried out at the end of the second school year.

Covariates

Many factors may influence second language acquisition (e.g., Spolsky, 1989). Some of these factors may be more influential than the teaching methods. To control for effects due to such factors, information on a number of covariates was collected.[5] For instance, information on child characteristics, such as language background, sex, socioeconomic status, behavioral problems, and IQ, was gathered through a sociolinguistic questionnaire that was assembled by the authors (see Appendix 2).

The sociolinguistic data were based on teacher judgments of the children's background and on information available in the pupils' files. To determine whether the two groups had different backgrounds, the sociolinguistic data were statistically tested with an ANOVA. The groups were found to differ significantly on two covariates.

First the professional level of the father, measured on a 7-point scale, indicated that the fathers of the children in the experimental treatment had jobs of a higher professional level ($M = 1.43$, $SD = 1.28$) than did the fathers of the children in the control treatment ($M = .69$, $SD = .97$), with $p < .02$.

Second, parents of children in the control treatment spoke their mother tongue with each other significantly more often ($M = .93$, $SD = .26$) than did parents of children in the experimental treatment ($M = .70$, $SD = .47$), with $p < .03$. Other covariates did not show significant group differences. The covariates on which the groups differed significantly had little effect on the dependent variables and therefore could be excluded.

A correlation matrix was calculated to study the effect of covariates on dependent variables. The number of covariates available for use in later analysis was found to be too high in relation to the number of subjects. On the basis of

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Huitema's inequality, as described in Stevens (1992), four covariates were acceptable for the multivariate analysis that was planned.[6] Therefore, only those covariates that correlated at least .40 with more than one dependent variable were incorporated in the analysis. The most salient covariates were two dummy variables, Turkish nationality and language spoken between parents and children is predominantly mother tongue, which originated from the nominal variables of nationality and language spoken between parents and children. Because they were dummy variables, their respective partners, Moroccan nationality and language spoken between parents and children is predominantly Dutch, had to be included as covariates as well. The third values for these nominal variables, other nationalities and language spoken between parents is both mother tongue and Dutch, did not have to be made explicit as they were implicitly coded. This amounted to four covariates being included in the MANOVA.

Treatments

The children in the experimental treatment worked through 20 formats (60 hours of format acting), and the children in the control treatment worked through both KNOOP and LAAT (60 hours of teaching). In both groups there were slight variations in how much teaching was done.

Apart from this, children in both conditions were taught in Dutch throughout each schoolday, and many other Dutch language activities were done with both the experimental and control groups. Children in both conditions were in similar situations and had comparable backgrounds, which was confirmed by a statistical analysis. Therefore, the most significant difference between the two groups of children was in teaching method.

Data collection procedures

Data collection on pretest language proficiency and intelligence was started in November 1993. Children were tested at their own schools by the first author. Before Christmas, all language data and data concerning IQ, hearing problems, and problem behavior were collected. The TRFs were filled out by the teachers in November and processed with the CBCL interpretation program on the computer. Midtest language proficiency data were collected in October 1994, this time by the first author with the help of a student. Finally, posttest language proficiency data were collected in June 1995, again by the first author with the help of six students, trained by the first author.

Statistical analysis

All data were entered into an SPSS database. To examine the hypothesis on the effects of the methods, MANOVA for repeated measures were done on the two sets of dependent variables. The first set involved three measures, and the second set involved two measures. To indicate effect sizes, eta squared[7] was used.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Data screening

Prior to data analysis all variables were examined using various SPSS programs for accuracy of data entry, missing values, and fit between their distributions and the assumptions of multivariate analysis. This was done separately for the 29 control group children and the 32 experimental group children.

Because of the small number of cases, it was not possible to remove all cases with missing data. Therefore, a rule was set: variables and covariates with more than 25% missing values over all cases were removed from the database, as were cases with more than 25% missing values over all variables and covariates. One case had to be removed from the experimental group, leaving 31 cases. Nominal covariates were recorded to dichotomous dummy variables.

All dependent variables were maintained because none reached the limit of 25% missing values over all cases. For five cases, missing values on dependent variables were replaced by the mean for all cases with the same nationality.

Four cases in the experimental group and six cases in the control group were identified as outliers on dependent variables. Because of the need to keep as many cases as possible, only one outlier on dependent variables was removed from the experimental group, leaving 30 cases; this was because it had extreme values on all dependent variables. A sensitivity analysis showed that the remaining outliers had an effect on the outcome of the analysis; therefore they were excluded from the final analysis.

Several dependent variables were severely skewed. This was especially the case at the first time of testing and less obvious at the third time of testing. This phenomenon is not surprising because the cases were selected from a group of immigrant children with little or no Dutch language proficiency, so one would expect a skewedness toward the low scores. During the experiment, both groups developed toward a more normally distributed population. This development was more or less the same in both the control and experimental groups. Because transformed data are not easy to interpret, an alternative procedure was followed to determine whether continuing with parametric statistics was still acceptable. For all dependent variables a parametric (student t test) and nonparametric (Kruskall–Wallis) comparison of group means was done. The outcomes of these tests were the same for each of the dependent variables, so it was decided to continue with the original plan for data analysis (i.e., MANOVA for repeated measures). No violation against the assumption of homogeneity of variance was detected.

RESULTS

Test results at the pretest

Table 1 shows the group means on the four dependent variables of Set I (sentence understanding, sentence building, receptive vocabulary, and active vocabulary), standard deviations, and spread in the data. The number of items is given under k. Group means were close and showed no significant differences between groups when subjected to an ANOVA. The proportion of correct items in the

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Table 1. *Subtest means and spread in pretest data, Set I*

	Experimental			Control			
Variable	*M*	*SD*	Range	*M*	*SD*	Range	*k*
Syntax							
Sentence understanding	12.6	10.4	0–38	11.9	8.8	0–33	64
Sentence building	2.5	5.7	0–30	1.7	4.1	0–15	46
Vocabulary							
Receptive	13.4	10.0	0–33	14.4	10.3	0–35	48
Productive	5.2	5.8	0–23	5.1	4.2	0–16	45

experimental group varied from 5.4% on the sentence building task to 27.9% on the receptive vocabulary task. For the control group, these proportions were 3.6% on the sentence building task and 30.0% on the receptive vocabulary task. Individual children sometimes performed considerably better or worse than group means (i.e., the starting levels of the selected children were not as low and homogeneous as had been hoped when designing the selection process; see range columns in Table 1).

Test results at the midtest

Test results of midtest data for the two sets of four dependent variables are described separately. The first set of dependent variables was measured at all three times of testing. This set consisted of the same variables as described in the test results for the pretest.

Both groups improved considerably on all four variables; the experimental group showed a mean difference of 1.8 standard deviations from the pretest; for the control group, the mean difference was 2.1 standard deviations from the pretest. The proportion of correct items in the experimental group varied from 26.2% on the productive vocabulary task to 63.2% on the receptive vocabulary task. For the control group, these proportions were 28.1% on the productive vocabulary task and 63.6% on the receptive vocabulary task. Table 2 shows the means, standard deviations, and spread in the data for these four tasks.

The second set of variables was measured at the midtest and the posttest only. Table 3 shows the means, standard deviations, and spread in the data for these tasks. Like the first set of variables, the means of both groups on the four tasks were close, but for the sound discrimination task, the group means differed significantly when subjected to an ANOVA, $F(1, 58) = 14.20$, $p = .00$, $\eta^2 = .20$. This difference was in favor of the experimental group.

The proportion of correct items in the experimental group varied from 15.3% on the morphology task to 79.4% on the sound discrimination task. For the control group, these proportions were 18.9% on the morphology task and 62.2% on the sound discrimination task.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Table 2. *Subtest means and spread in midtest data, Set I*

	Experimental			Control			
Variable	*M*	*SD*	Range	*M*	*SD*	Range	*k*
Syntax							
Sentence understanding	33.4	10.6	16–55	33.3	10.7	5–53	64
Sentence building	16.3	12.8	0–40	13.1	10.4	0–35	46
Vocabulary							
Receptive	30.3	10.5	3–46	30.6	10.5	1–42	48
Productive	11.8	7.0	2–29	12.7	7.0	1–26	45

Table 3. *Subtest means and spread in midtest data, Set II*

	Experimental			Control			
Subtest	*M*	*SD*	Range	*M*	*SD*	Range	*k*
Phonology							
Discrimination	39.7	7.0	23–49	31.1	10.3	10–46	50
Articulation	36.4	4.4	11–44	37.9	5.9	21–44	44
Morphology	5.5	5.8	0–22	6.8	6.2	0–20	36
Comprehension of text	12.0	4.4	6–21	10.6	5.6	0–19	24

Test results at the posttest

On the first set of dependent variables, both groups again improved. The experimental group showed a mean difference of 0.8 standard deviations from the midtest; for the control group, the mean difference was 0.9 standard deviations from the midtest. The proportion of correct items in the experimental group varied from 44.2% on the productive vocabulary task to 75.7% on the receptive vocabulary task. For the control group these proportions were 49.0% on the productive vocabulary task and 77.2% on the receptive vocabulary task. Table 4 shows the means, standard deviations, and spread in the data for these four tasks.

On the variables measured at the midtest and posttest only (Set II), the progress of both groups was smaller than it was for Set I. On the sound discrimination and sound articulation tasks, the reason for this is that the scores of the children on the midtest were already high. On the morphology and comprehension of texts tasks, the reason can be found in the difficulty of the tasks with regard to the age of the children. The experimental group showed a mean difference of 0.7 standard deviations from the midtest; for the control group, the mean difference was 0.6 standard deviations from the midtest. The proportion of correct items in the experimental group varied from 30.6% on the morphology task to 95.5% on the sound discrimination task. For the control group these propor-

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Table 4. *Subtest means and spread in posttest data, Set I*

	Experimental			Control			
Variable	*M*	*SD*	Range	*M*	*SD*	Range	*k*
Syntax							
Sentence understanding	43.3	8.8	21–60	41.3	7.8	26–55	64
Sentence building	21.6	12.2	1–41	21.7	10.1	0–39	46
Vocabulary							
Receptive	36.3	7.6	12–47	37.0	8.7	11–46	48
Productive	19.9	9.8	6–38	22.0	8.7	2–33	45

Table 5. *Subtest means and spread in posttest data, Set II*

	Experimental			Control			
Subtest	*M*	*SD*	Range	*M*	*SD*	Range	*k*
Phonology							
Discrimination	42.4	5.6	28–50	37.5	9.3	7–48	50
Articulation	42.0	3.6	26–44	40.2	2.9	32–44	44
Morphology	11.0	7.4	0–23	12.0	7.3	0–24	36
Comprehension of text	14.8	5.2	4–24	13.3	5.3	0–21	24

tions were 33.2% on the morphology task and 91.3% on the sound discrimination task. Table 5 shows the means, standard deviations, and spread in the data for these four tasks.

MANOVA results showed that, although the experimental group clearly outperformed the control group on the sound discrimination task (ANOVA results: $F(1, 58) = 346.78$, $p = .02$, $\eta^2 = 10.5$), the mean of the children in the control group increased much more than the mean of the children in the experimental group.

Examples of progression

To illustrate these numbers, some examples of linguistic production by the children are given. As is already clear from these results, the starting levels were not as low for all children as was aimed at in the selection of subjects. Some children were already speaking quite a lot of Dutch, while others did not speak or understand a word. The following examples show the differences in the utterances between a child who is a real beginner (Child1) and a child who is more proficient (Child2).

At the beginning of the project the children with little Dutch language proficiency communicate their wishes and thoughts through one-word or two-word sentences.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Teacher: Wat doet het meisje? (What is the girl doing?)
Child1: Spelen. (Play.)
[When Ike saw a girl playing with a football.]
Teacher: Ja ze is aan het spelen hè? (Yes, she is playing, isn't she?)
Child1: ⟨Bal⟩ voetbal. (⟨Ball⟩ football.)

In the last utterance this child specifies the kind of ball the girl is playing with. In another example of Child1 we see a two-word sentence.

Teacher: Wie weet waar hij moet beginnen. (Who knows where to begin?)
[Which pictures comes first in sequencing the story.]
Child1: Weet ikke. (I know.)

In this example the teacher's question is already quite complex, yet Child1 seems to understand, as his response is appropriate, though incomplete. The English translation is a complete sentence, but in Dutch it is not; a direct object is omitted.

Towards the end of the project we see more complex utterances, and we even find Child1 correcting another child.

Child: Marja, gaan we die spelletje maken. (Marja, are we going to make that game.)
Child1: Maken, jij zegt maken. (Make, you say make.)
Child1: Niet maken. (Not make.)

When we compare Child1 with Child2, a boy who already speaks quite a lot of Dutch at the beginning of the project, we find utterances like the following:

Playmate: Je mag niet aankomen. (You may not touch.)
[The microphone that he is wearing.]
Child2: Ja, jij ⟨komt ook⟩ aankomen. (Yes, you ⟨also touch⟩ touching.)

The utterance is more like a sentence, although Child2 gets lost in the syntax of the utterance. The sentence begins fine, but he does not know how to continue and tries to correct himself. Unfortunately, this correction is not right.

Child2: Jessin heb dat boven gedaan. (Jessin have done that upstairs.)
[When he sees a picture of a boy bouncing a ball. Upstairs is the playroom.]

This sentence is already an amplified sentence; all necessary arguments (*Jessin, that*) and a predicate (*have done*) are present, as well as a facultative argument (*upstairs*), although the predicate is not correct.

Toward the end of the project we find even more amplified sentences from Child2, although there are still imperfections in word order.

Child2: Je kan lekker ook in het zand spelen. (You can play nicely also in the sand.)
[When the teacher asks him what one can do for fun.]

Sometimes the vocabulary of Child2 is insufficient as well. Like many children in the project, Child2 is very creative in inventing new words.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Teacher: Eerst doe je je onderbroek aan en dan . . . (First you put on your knickers on then . . .)
Child2: Je binnenhemd. (Your "innershirt.")

On the whole Child2 is quite fluent at the end of the project. He makes complex amplified sentences, can tell a short story, and is capable of logical reasoning in Dutch.

Teacher: En de reus zegt, "ik heb honger, mijn buik is niet gevuld." (And the giant says, "I am hungry, my stomach is not filled.")
Child2: Dan is je buik leeg. (Then your stomach is empty.)

From these examples it can be concluded that within-group differences in language proficiency between children are clear, not only in test data but also in more spontaneous language production.

MANOVA procedures and outcomes

To search for significant differences between group means, MANOVA for repeated measures were done. These were done separately for the two sets of dependent variables. The first set (Set I) consisted of the subtests (a) receptive vocabulary, (b) productive vocabulary, (c) sentence understanding, and (d) sentence building, which was measured three times during the project. The second set (Set II), which was measured only at the midtest and posttest, consisted of (a) sound discrimination, (b) sound articulation, (c) morphology, and (d) comprehension of texts.

Results for Set I. An SPSS MANOVA was used to perform a multivariate analysis with repeated measures. Six cases with extreme values on sentence building were excluded from the analysis. Using Wilks's criterion, there was a multivariate effect of the (dummy) covariates, nationality and language spoken between parents and children, on the dependent variables of Set I, $F(16, 135) = 2.55$, $p = .01$, $\eta^2 = .65$. In the univariate analyses, the (dummy) covariates, Turkish nationality and language spoken between parents and children is predominantly mother tongue, had a significant negative effect on all four dependent variables of Set I. The variable, receptive vocabulary, was negatively influenced by the language covariate, nationality.

No significant differences were found between the two conditions, $F(4, 44) = 1.26$, $p = .29$. The profiles of the two groups did not deviate significantly from parallelism, $F(8, 44) = .87$, $p = .58$ (Wilks's criterion), so they may be considered to be parallel. The deviation from flatness was significant in a multivariate analysis, $F(8, 44) = 110.82$, $p = .00$, $\eta^2 = .95$. This was confirmed in the univariate analyses. The within-subjects factor, time, was significant for all dependent variables, but interaction effects of time by condition were not. Although both groups showed a significant increase in means on all dependent variables, there were no effects of the variable, condition, nor any interaction effects of time and condition.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Results for Set II. For Set II, there was a multivariate effect of the covariates on the dependent variables, $F(16, 144) = 3.42$, $p = .00$, $\eta^2 = .63$. In the univariate analyses the covariates showed a significant negative effect on the dependent variables, comprehension of texts and morphology. There were no significant effects of covariates on the sound discrimination and sound articulation tasks.

In the multivariate analysis, there was a significant effect of condition, $F(4, 47) = 4.20$, $p = .005$, $\eta^2 = .26$, but in this case the profiles of the two groups deviated significantly from parallelism, $F(4, 51) = 3.70$, $p = .01$, $\eta^2 = .33$ (Wilks's criterion). The deviation from flatness was also significant, $F(4, 51) = 29.27$, $p = .00$, $\eta^2 = .70$. In univariate analyses, two subtests, sound articulation and sound discrimination, showed a significant difference between the two groups. Again, both results fail to benefit one condition. Sound articulation was in favor of the children in the experimental condition, whereas sound discrimination was in favor of the children in the control group.

In other words, both groups showed a significant increase in means on all dependent variables of Set II over time. There was a significant multivariate effect of condition, but the univariate tests demonstrated that these effects did not favor one condition.

DISCUSSION

We may conclude that the effects of the methods involved in this research were hardly different following two years of teaching. Two significant group differences that were found in the analysis of Set II, sound discrimination and sound articulation, were in favor of the control condition and the experimental condition, respectively. The results of the six other dependent variables were not significantly different between the groups.

We are not completely sure why we obtained these results. The skills that show no difference between the two conditions (receptive and productive vocabulary, sentence building and understanding, morphology, and comprehension of texts) may need more intensive training to invoke significant differences between the two conditions than either of the methods included in this research could account for. The two differences that were found were within tasks that cover the same domain (phonology), whereby the control group improved more on a passive task and the experimental group improved more on a productive task. The fact that the experimental group improved more on the articulation task but not on the discrimination task is quite puzzling. However, it is important to keep in mind that the children in the experimental condition outperformed the children in the control condition on both the sound discrimination task and the sound articulation task. One would expect that, in order to reproduce a word that is read aloud by the tester, one would have to be capable of sound discrimination as well as sound articulation, which is the case indeed if we look at the raw data. Children in the experimental condition did perform better on both tasks.

It is most likely that the results of the growth are due to implicit training effects as a result of the structure of the respective methods. The test procedure used during the sound articulation task, in which the children had to reproduce

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

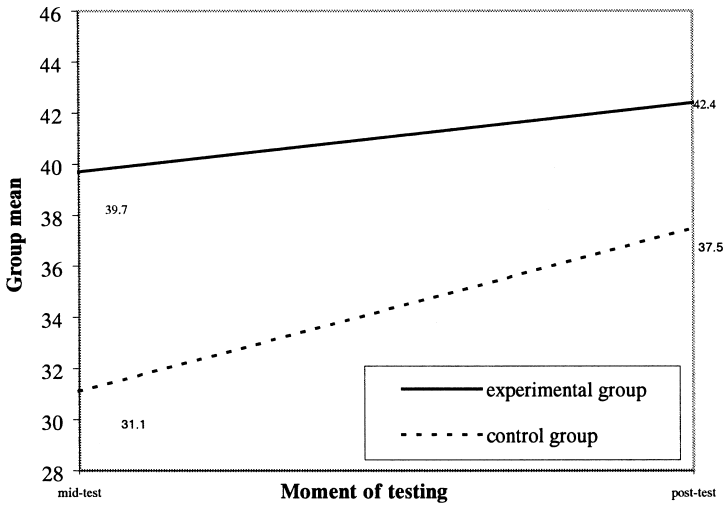


Figure 1. Growth in sound discrimination.

words that were read aloud by the tester, leads to the thought that the children in the experimental condition were trained in this skill, because the procedure during the format play, whereby the children repeat a sentence after their teacher, resembles the test procedure. Similarly, we can say that the children in the control condition performed better on the sound discrimination task, which is a receptive task, because the emphasis in KNOOP and LAAT is on receptive tasks. Here, the children spent a considerable amount of time listening to the teacher, even during the lessons with LAAT, where productive vocabulary and speech proficiency received more attention.

Looking at Figures 1 and 2, this effect of training becomes more clear. In Figure 1, the slope of the line of the control group is much steeper than the slope of the line of the experimental group. In Figure 2, we find the opposite picture. The slope of the line of the experimental group is much steeper than the slope of the line of the control group. When these four lines are displayed in one picture (Figure 3), we find a striking parallism between the slopes of skills that were (implicitly) trained through the structure of a method and those that were not trained. The flatter slopes seem to demonstrate a natural growth in means on a task that was not trained, and the steeper slopes seem to demonstrate the growth in means of a trained skill.

Although these results seem obvious, we are cautious with these suggestions, especially because the findings are not supported by the results on the other tasks. The results of both groups in receptive vocabulary, which is explicitly trained in the KNOOP and LAAT methods, are equal. The same goes for productive vocabulary: the experimental group did not perform better on this task, although it was trained (implicitly) during the lessons with the Dinocroc method. The effects of training are not present here. The skills that showed no difference

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

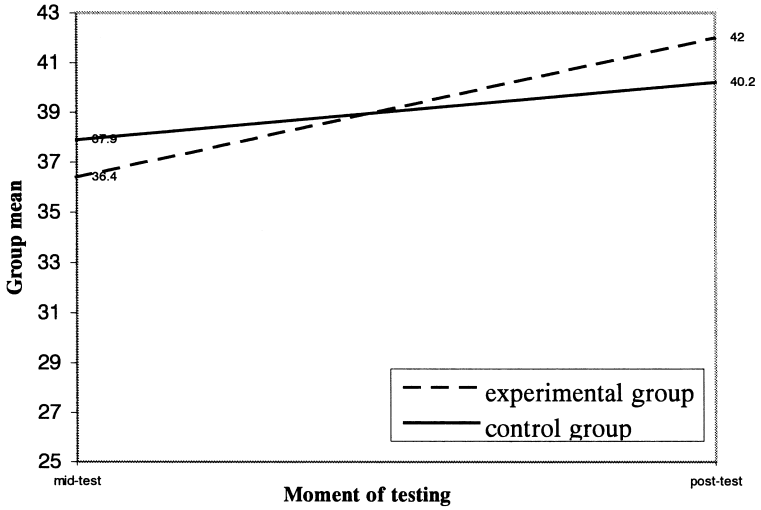


Figure 2. Growth in sound articulation.

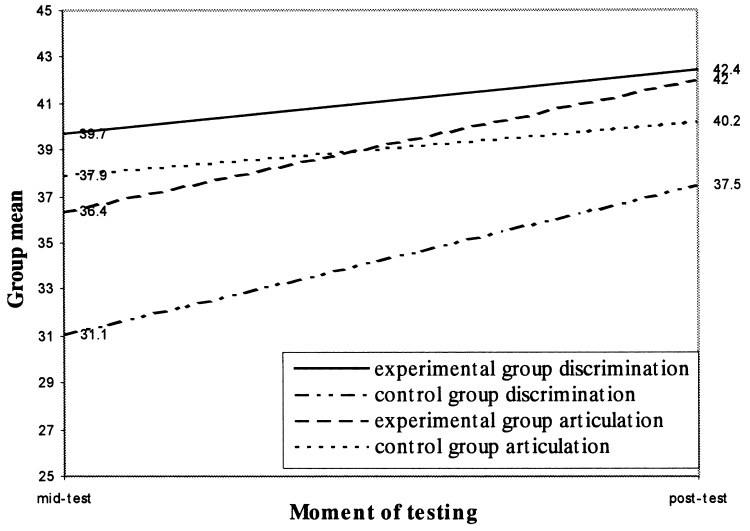


Figure 3. Parallelism of slopes in trained task and natural growth.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

between the two conditions (receptive and productive vocabulary, sentence building and understanding, morphology, and comprehension of texts) may need more intensive training to invoke significant differences between the two conditions or they may be less susceptible to classroom training in general.

Another reason to be cautious with these findings is that the skills of sound discrimination and sound articulation were measured only at the midtest and posttest. At the pretest we could not measure them due to the lack of an appropriate test, whereas at the midtest the children had already been taught with their respective methods throughout that year. Therefore, we do not know what happened during the first year. At what level did the groups start? With respect to the sound discrimination task, the following thoughts arise. Had there already been in the first year a growth leap in the score of the experimental group? Did the control group develop in the same way during the first year or did the development in the first year show a flatter slope? Was the mean of the experimental group restricted because of the maximum score that is possible on this task and were the scores of the children in the experimental group already more in the direction of the maximum than the scores of the children in the control group?

With respect to the sound articulation task, similar questions arise. What happened during the first year? Did the children in the control group go through a growth leap? Did the children in the experimental group develop in the same way during the first year or did their development show a flatter slope? Again, was the mean of the experimental group restricted because of a maximum number of items? On this task there were even more experimental group children who reached the maximum score than on the sound discrimination task. On the whole, it is very unfortunate that we have no data available on the group means of the two groups at the first time of testing.

With regard to the validity of the present study some remarks need to be made. First, how reliable is the implementation of the treatments? Teachers in the experimental classes were, in a way, forced to stick to the prescriptions because they were video taped twice a month and were observed directly in the classroom by the internal counselors during one of these sessions.[8] Apart from this, the fact that they were given information frequently and had experienced an intensive training on how to work with the method at the beginning of the project year should have contributed to a more reliable implementation. The presence of individual differences cannot be ruled out, of course. In the control classes there was less control on the implementation of the treatment. There were no observations or video sessions, and the teachers were working with methods they had used for some time. Because of this, their individual adaptations to the programs may have drifted away more from the original prescriptions. Thus how the treatment was implemented probably varied more among teachers of the control classes.

The second remark is strongly related to the first. Can the finding that the Dinocroc method is equally effective be attributed to the fact that teachers in the experimental condition had been working with more dedication than the teachers in the control condition? This question cannot be answered, but possible positive effects of this kind may have a countereffect in the fact that some

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

teachers were working with an unfamiliar teaching method which demanded unusual skills.

We used standardized language tests, which assess language proficiency in general rather than the level of mastery of the contents offered by a specific teaching method. If program-bound tests had been used, we could have studied what it was that children had learned from their respective methods. But in order to answer the hypothesis properly, we had to use tests that could be compared between the two treatments, and therefore the tests had to be program-independent. It may never be possible to resolve the conflict between testing what has been taught and testing so that results can be compared between experimental and control groups. To see if children would perform better on method-bound tests, we compared the two groups on the items that were explicitly taught in the respective methods and coincidently present in the vocabulary tests. For the productive vocabulary test, nine items were present in the vocabulary test and in KNOOP and LAAT. For each case the mean score on these items was calculated. The group means of the two conditions on these nine items were compared for all three times of testing, but in both multivariate and univariate tests no significant differences were found. Only two items of the productive vocabulary test were explicitly taught in the Dinocroc method; although this may be too little to make a proper analysis, group means on these items were compared. Again, no significant differences between the two groups in their mean score on these items were found. For the receptive vocabulary tests, similar analyses were done: ten items of the receptive vocabulary test were explicitly taught with the Dinocroc method, and six items were explicitly taught with KNOOP and LAAT. In both multivariate and univariate analysis, the groups showed no significant differences in their mean scores on these items. These results show that, at least for the vocabulary tests, the children did not perform better on method-bound items. This seems to justify our choice for a standardized language assessment.

Apart from this, discrete test data are not suitable to assess all aspects of language proficiency, especially those that are the focus of the format-based method. Testing communicative skills, for instance, was not included in this study, but relevant data were collected and will be studied in the future. For these reasons, no final conclusions can be drawn about the impact of the format-based and the code-oriented methods on the basis of the results presented here.

After considering these threats to the validity of the present study, we may conclude that after two years of training the two treatments scarcely differed, and that none of the treatments appeared to be biased during the research.

Besides its linguistic impact, which needs more research, it is clear that the format-based method is equally effective as KNOOP and LAAT, but that the didactic approach of the format-based method fits the specific needs of kindergarten children better than the code-oriented method. All participants frequently reported in their log books that both children and teachers worked with this method with great pleasure. From these log books a hypothesis for further research arose. Children in the experimental treatment were reported to be much more communicative and more eager to participate verbally during classroom activities when teachers compared them to children they taught in the years

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

before they used the format-based method. This reported difference in attitude of the children may have implications for their future language acquisition. Damhuis (1995) reviewed the literature on communication and interaction in second language acquisition and concluded that participation in interaction is beneficial to second language acquisition. This effect is even stronger when children take the initiative instead of playing a responsive role. In this light, the second language acquisition of the children in the experimental treatment would be likely to benefit from their enhanced positive attitude, if given more time. Apart from the research on communicative skills, future research within the European project in the Netherlands will be dedicated to this hypothesis.

GENERAL DISCUSSION

It is interesting to compare the research findings of the Dutch experiment, where the Dinocroc method is used as a second language teaching method in kindergarten (ages 4 to 6 or 7), with the results of the Basque experiment (Azpillaga et al., 2001), where the method is used as a foreign language teaching method in primary school (ages 8 or 9 to 10 or 11). In the latter experiment, the experimental group outperformed the control group, whereas in the Dutch experiment the research findings showed small differences between the experimental and control groups but not unambiguously in favor of one condition. Which factors can account for the differences in these research outcomes? What are the differences between foreign and second language learning and can these differences explain research outcomes? An attempt will be made here to answer these questions, but it is imperative to keep in mind the fact that we are not making a comparison of the actual levels of achievement between the Basque and the Dutch research groups.

In general, studies concerning the rate of acquisition have concluded that older is better. Krashen, Long, and Scarcella (1979) concluded that older learners learn more rapidly than younger learners in terms of rate of acquisition. The study of Snow and Hoefnagel-Höhle (1978) supported these findings. Their study investigated the naturalistic acquisition of Dutch by 8- to 10-year-old English children, 12- to 15-year-old adolescents, and adults over a 10-month period. With regard to morphology and syntax, the adolescents did best, followed by the adults and then the children. There were only small differences in pronunciation, and the grammar differences diminished over time as the children began to catch up.

The effect of age on second language achievement seems to be the opposite: younger is better, although there is a difference between formal learning environments and naturalistic learning situations. In formal learning environments, the conclusions do not support the hypothesis that younger is better. Burstall (1975) reported on a pilot scheme in England and Wales. She compared two groups of students with five years of instruction. One group had begun learning French at age 8, whereas the other had begun at the beginning of secondary school (age 11). She found that the older learners were "consistently superior." Harley (1986) compared the levels of attainment of children in French bilingual programs in Canada. She compared early and late immersion students after

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

1,000 hours of instruction. She found that older students demonstrated a greater overall control in the acquisition of the verb system. Although the early immersion group demonstrated higher levels of attainment at the end of their schooling, this result may reflect the additional number of years of instruction they had received rather than their starting age. Of course, none of the studies mentioned here show the effects of age on learners' ultimate level of attainment. A possible explanation (Singleton, 1989) is that formal learning environments do not provide learners with the amount of exposure needed for the age advantage of young learners to emerge.

Studies of learners in naturalistic learning situations have provided evidence that younger is better. Learners who start as children achieve a more native-like accent than do those who start as adolescents or adults (Asher & Garcia, 1969; Oyama, 1976; Tahta, Wood, & Loewenthal, 1981). Similar results have been obtained for the acquisition of grammar (Johnson & Newport, 1989; Patkowski, 1990). "Concerning the hypothesis that those who begin learning a second language in childhood in the long run generally achieve higher levels of proficiency than those who begin in later life," Singleton (1989, p. 137) concluded that there is good supportive evidence and that no actual counter evidence exists.

We can say that older learners have an initial advantage where rate of learning is concerned. They will eventually be overtaken by child learners who receive enough exposure to the second language. This is less likely to happen in instructional than in naturalistic settings because the critical amount of exposure is usually not available in the former.

Taking all this into account, the differences in effect of the method that were found in our study and in Azpillaga et al. (2001) still cannot be explained by the differences in the ages of the research groups, as both studies used control groups of similar age. Therefore, age can hardly account for the different outcomes between the studies. One would expect to see an influence of age on research outcome only if the groups compared within each of the studies showed a significant age difference.

It might be that older children (like those in the Basque study; Azpillaga et al., 2001), when provided with an optimal learning environment (as in the Dinocroc method), learn more and do so more rapidly than younger children when they are brought into an optimal language environment. Further explanations of differences in research outcomes might be found in the fact that in the Basque experiment the children came from higher social economic classes than did those in the Dutch study. A combination of age with a positive language environment and a higher social economic status of the learners might explain the higher acquisition rates in the Basque experimental groups, although this hypothesis is highly speculative.

The most parsimonious explanation of the fact that children in the Dutch experimental group, in contrast with those in the Basque experiment (Azpillaga et al., 2001), did not achieve a higher level of language acquisition in comparison with the control group is that the amount of learning Dutch as a second language with a specific method was only a very small amount of all the second language learning that took place during the schoolday. It can be stated that

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

children in both the control and the experimental conditions in the Dutch experiment had more in common, and that it is therefore impossible to find differences in language achievement between these two groups. Both teaching methods used in the Dutch research only account for 60 hours of second language education, divided over two years of teaching. The language learning opportunities that the children had during the rest of their time in school were likely to be quite comparable. In the Netherlands, children start school at the age of 4. During the first two years of kindergarten, children spend 880 hours per year at school; this amount is fixed by law. This means that only 60 hours (3%) of the 1,760 teaching hours children receive in their first two years at school are different, and that the other 1,700 hours (97%) provide them with more or less the same opportunities for language learning. Moreover, the two methods used as treatments in this study share certain features. What happens during the many hours of life outside school during these two years is even more likely to be comparable. Many studies report the influence of, for instance, socioeconomic status, which has more influence on second language acquisition than any other factor. Since the children in the experimental and control groups had generally the same socioeconomic background, the hypothesis that 60 hours of second language lessons would make a difference, even when the teaching techniques would be substantially different, is obviously not viable.

In the Basque study (Azpillaga et al., 2001) the opportunities for children to learn a foreign language outside the school or outside foreign language lessons rarely existed. Introducing two very different methods in this setting – one highly comparative with natural second language acquisition and one more of an instructional method – could thus be assumed to lead to differences in language achievement between the two groups.

APPENDIX 1

EXAMPLE ITEMS OF THE SUBTESTS USED IN SET I

Subtest one and two, Sentence Understanding (TAK): Variable sentence understanding

Two of the three sentence understanding tasks of the TAK were used. Each tests a slightly different aspect of sentence understanding. Each test consists of 32 items. The test leader shows a set of three pictures, and the child has to point out the appropriate picture. Officially there is no break-off criterion, but we restructured the test and did use one.

(1) Sentence patterns: The tree is higher than the house.
afbeelding

(2) Relations in meanings: The girl forgot to close the door.
afbeelding

The results of these two subtests are added and can amount to a maximum of 64 items.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Subtest three, Sentence Building (TAK): Variable sentence building

The test leader reads aloud a long and complex sentence; the child has to repeat this sentence. With respect to the function words and the sentence patterns in each sentence, the child's reproduction has to be perfect. The idea is that the child can only repeat the sentence correctly if he or she understands what is meant. This subtest consists of 22 function words and 24 sentence patterns, which amount to a maximum score of 46 items. Again there is officially no break-off criterion, but as with the sentence understanding tasks we did use a break-off criterion on the restructured task.

(1) The teacher *will* give him a new paper *that is clean.*
(2) *What you have to do* is listen carefully *if* someone is talking.

Subtest four, Productive Vocabulary (Taalplan Kleuters): Active vocabulary

The child is shown a picture of a lexical item and has to say what is shown. The test consists of 45 items. The testing is terminated when five successive items are named incorrectly.

This is a_____
afbeelding

Subtest five, Receptive Vocabulary (Taalplan Kleuters): Variable receptive vocabulary

The child is shown a set of four pictures. The test leader says a word, and the child has to point out the right picture. The total number of items is 48. The test is ended when only two out of eight successive items are identified correctly.

Where do you see an umbrella?
afbeelding

EXAMPLE ITEMS OF THE SUBTESTS USED IN SET II

Subtest one, Discrimination of Sounds (TAK): Variable discrimination of sounds

In this subtest the test leader reads aloud 50 pairs of real words. These sets can be either the same words or different words. In cases where the set consists of different words, the phonological resemblance of the words is high. The child then has to decide whether the two words are the same.

1. Boer – Buur (Farmer – Neighbor)
2. Vel – Veel (Skin – A lot)
3. Kan – Kan (Jug – Jug)

These test items are translated to give an idea of the difference in meaning between Dutch words that have only small phonological differences. One can imagine these kind of words in any language. There is no break-off criterion.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Subtest two, Articulation of Sounds (TAK): Variable articulation of sounds

In this subtest the test leader reads a word aloud. The child has to repeat this word aloud. The subtest consists of 44 items. There are three categories of words: (1) consonant–vowel–consonant (C-V-C) words, (2) C-C-V-C words, and (3) C-V-C-C words. There is no break-off criterion.

1. Rug (C-V-C) (Back)
2. Knop (C-C-V-C) (Button)
3. Koorts (C-V-C-C) (Fever)

Subtest three, Morphology (TAK): Variable morphology

Four different grammatical proficiencies are tested. For each item, pictures are available to trigger the child's response.

(1) plural, 12 items:
Een paard, twee _____? One horse, two _____?
afbeelding

(2) past perfect, 9 items:

Rosita is een bal aan het gooien.	Rosita is throwing a ball.
Wat heeft Rosita gedaan?	What did Rosita do?
Ze heeft de bal _____?	She has (thrown) the ball.

afbeelding

(This kind of elicitation is only possible in Dutch because the past perfect is situated at the end of a sentence.)

(3) past, 9 items:

Ik zag Kees lopen.	I saw Kees walking.
Hij riep 'help' toen hij _____?	He shouted "help" when he _____?

afbeelding

(4) comparative-superlative, 6 items:

Dit is een grote boom.	This is a big tree.
Maar deze boom is véél _____?	But this tree is much _____?
En deze boom is de aller _____?	And this tree is the _____?

afbeelding

There is no break-off criterion.

Subtest four, Comprehension of Texts: Variable comprehension of texts

The test leader reads aloud a short story. Afterward, the child has to answer questions about the story. The test consists of six stories, each with four questions.

Story one
It is almost six 'o clock.
Mother sees that she is out of bread.
There is nothing to eat for tonight.
She calls Ali.
He must go to the baker shop and buy a loaf of bread.
Ali goes to the baker shop.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Then he meets a friend.
Who asks if he wants to come and play.
Together they go to the park.
Ali has totally forgotten the bread.

Questions

1. What did Ali have to buy?
2. Where does Ali go with his friend?
3. Did Ali buy a loaf of bread?
4. Do they have something to eat for tonight?

APPENDIX 2

SOCIOLINGUISTIC CONTEXT QUESTIONNAIRE

Name: Sex:
Date of birth: Place of birth:
In the Netherlands since:*
School:
Profession of father: Nationality of father:
Education of father: In the Netherlands since:
Profession of mother: Nationality of mother:
Education of mother: In the Netherlands since:
Number of brothers older than the child:
Number of sisters older than the child:
Number of brothers younger than the child:
Number of sisters younger than the child:
Family structure:
Nuclear family / Extended family / Single parent family
Language spoken between parents:
mainly Dutch / mainly mother tongue / both
Language spoken between siblings:
mainly Dutch / mainly mother tongue / both
Language spoken between parents and children:
mainly Dutch / mainly mother tongue / both
Dutch language proficiency of mother: Low / Average / Good
Dutch language proficiency of father: Low / Average / Good
Dutch language proficiency of siblings: Low / Average / Good
Contacts of the family with Dutch speaking people: None / Some / A lot
Religious orientation: Christian / Islam / Other / None
Cultural orientation: Dutch culture / Culture of mother country
Child participated in preschool intervention programs:
No / Yes: Which
Child currently participates in intervention programs:
No / Yes: Which
IQ as measured with the KSSON:
Score on internalizing problem behavior as measured with the TRF:
Score on externalizing problem behavior as measured with the TRF:

*if the child was not born in the Netherlands

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

ACKNOWLEDGMENTS

We would like to thank those who made the project and data collection possible:

All the teachers, internal counselors, headmasters and school boards, children and parents of the schools involved in the experimental treatment – De Flevoparkschool, De Knotwilg, and De Joop Westerweel school – and in the control treatment – De Jan Pieterszoon Coen school, Het Kruispunt, De Roos, and De Rosa Boekdrukkerschool – for their participation and hard work and for opening the classroom doors to research. We especially want to thank the teachers and internal counselors of the experimental schools who offered their spare time and parts of their summer vacation to attend meetings and training sessions. We thank the school boards for the extra salaries they paid in order to give the participating teachers and facilitators time off to spend on the project.

The European Community's Bureau Lingua in Brussels for several grants that made the European project possible: Lingua Program projects numbers 92-01/0524/I-IB, 93-01/0863/I-IB, 94-01/1432/I-IB, 92-09/0783/I-VB, 93-09/1286/I-IB, 94-09/1786/I-IB.

The Dutch Ministry of Education and Science, De projectgroep NT2: a 1-year grant made it possible to buy necessary equipment.

All the students who helped with data collection and data entry, especially Esther Windhorst and Marieke van Heerde.

Peter de Jong of the Free University of Amsterdam, Faculty of Psychological and Educational Sciences, for his permission to let us use the Toetsen bij Woordenlijst 4- tot 6-jarigen.

And finally, Dr. J. Hox and Dr. G. van den Wittenboer for their methodological advice.

NOTES

1. In the Netherlands kindergarten covers the first two years of primary education. Children enter kindergarten at age 4.
2. The method still requires a better name.
3. This selection was carried out to enable the study of the influence of behavioral variables of the children in the process of language acquisition. This aspect of the research is not included in this article.
4. As for the vocabulary domain, the Toetsen Woordenlijst voor 4- tot 6-jarigen were chosen; these were derived from tests (Toetsen Taalplan Kleuters; Damhuis & Schoonen, 1993) developed for the Taalplan Kleuters method (Boers, Kienstra, & Padmos, 1993), which were available for research purposes. The Taalplan tests are method-bound, but two method-independent vocabulary tests (one receptive and one productive) were assembled by selecting items from the Toetsen Taalplan Kleuters and by rank-ordering these items according to a vocabulary list (Woordenlijst voor 4- tot 6-jarigen; Damhuis, de Glopper, Boers, & Kienstra, 1992). This list is based on teacher judgment of the desirable point at which words are to be mastered.
5. Information on the following covariates was not yet available but will be included in the future: the intensity of the second language teaching program, as assessed through teacher logs, and school characteristics, such as total number of pupils, number of classes, school policy, school climate, school organization, the various kinds of training and support for the teachers, and pedagogical conditions, as assessed through a school policy questionnaire.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

6. Huitema's formula to calculate the number of covariates that is allowed in a multivariate analysis in relation to the number of cases is as follows: $[C + (J - 1)] / N < .10$, where C = number of covariates, J = number of groups or levels of a factor, and N = number of cases. In our case, the result of this formula is $[4 + (2 - 1)] / 59 = .08$.
7. The eta squared used to measure the strength of the association in the repeated measures MANOVA is not the standard η^2, but $\eta^2 = 1$ (Tabachnick & Fidell, 1989). When ANOVA results are reported, the standard η^2 (explained SS / total SS) is used.
8. It is in the light of the study of the influence of behavioral variables of teachers that the video taping and direct observations were done.

REFERENCES

Achenbach, T. (1988). *Teacher Report Form*. Burlington: University of Vermont, Department of Psychiatry.

Appel, R. (1984). *Immigrant children learning Dutch: Sociolinguistic and psycholinguistic aspects of second-language acquisition*. Dordrecht: Foris

Appel, R., & Vermeer, A. (1994). *Tweede-taal verwerving en Tweede-taalonderwijs*. Bussum: Coutinho.

Asher, J. (1969). The total physical response approach to second language learning. *Modern Language Journal, 53,* 13–7.

Asher, J., & Garcia, R. (1969). The optimal age to learn a foreign language. *Modern Language Journal, 53*, 334–341.

Azpillaga, B., Arzamendi, J., Etxeberria, F., Garagorri, X., Lindsay, D., & Joaristi, L. (2001). Preliminary findings of a format-based foreign language teaching method for school children in the Basque Country. *Applied Psycholinguistics, 22,* 35–44.

Boers, M., Kienstra, M., & Padmos, M. (1993). *Taalplan Kleuters*. Rotterdam: Het projectbureau OVB Rotterdam.

Boogaard, M., Damhuis, R., de Glopper, K, & van den Bergh, H. (1990). *De Nederlandse taalvaardigheid van allochtone en Nederlandse Kleuters. Peiling van de taalvaardigheid in het Nederlands van Surinaamse, Antilliaanse, Marokkaanse, Turkse en Nederlandse leerlingen aan het einde van de kleuterperiode*. Amsterdam: Swets & Zeitlinger.

Bruner, J. (1975). From communication to language: A psychological perspective. *Cognition, 3,* 225–287.

(1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Burstall, C. (1975). Factors affecting foreign-language learning: A consideration of some relevant research findings. *Language Teaching and Linguistic Abstracts, 8,* 105–125.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation*. Chicago: Rand McNally.

Damhuis, R. (1995). *Interaction and second language acquisition*. Amsterdam: IFOTT & SCO-Kohnstamm Instituut.

Damhuis, R., de Glopper, K., Boers, M., & Kienstra, M. (1992). *Woordenlijst voor 4- tot 6-jarigen*. Rotterdam: Het Projectbureau OVB Rotterdam.

Damhuis, R., & Schoonen, R. (1993). *Taalplan kleuters: Toetsen*. Rotterdam: Het projectbureau OVB Rotterdam.

Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon.

Erdfelder, E., & Faul, F. (1996). *Gpower: A general power analysis program*. Version 2.0. Freeware.

Harley, B. (1986). *Age in second language acquisition*. Clevedon: Multilingual Matters.

Johnson, L., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Kohnstamm, G., Schaerlaekens, A., de Vries, A., Akkerhuis, G., & Froonincksx, M. (1981). *Nieuwe streeflijst woordenschat voor 6 jarigen*. Lisse: Swets & Zeitlinger.

Krashen, S., Long, M., & Scarcella, R. (1979). Age, rate and eventual attainment in second language acquisition and language learning. *TESOL Quarterly, 13,* 573–582.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Trowbridge: Redwood Burn.

MacWhinney, B. (1991). *The Childes Project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

Ochs, E., & Schieffelin, B. (1995). The impact of language socialization on grammatical development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 73–93). Cambridge: Basil Blackwell.

Oyama, S. (1976). A sensitive period in the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research, 5,* 261–285.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emir Flege. *Applied Linguistics, 11,* 73–89.

Singleton, D. (1989). *Language acquisition: The age factor*. Clevedon: Multilingual Matters.

Snijders, J., & Snijders-Oomen, N. (1975). *Snijders-Oomen Niet Verbale Intelligentieschaal*. Groningen: Wolters-Noordhoff.

Snow, C., & Hoefnagel-Höhle, M. (1978). The critical age for language acquisition: Evidence from second language acquisition. *Child Development, 49,* 1114–1128.

Spolsky, B. (1989). *Conditions for second language learning. Introduction to a general theory*. Oxford: Oxford University Press.

Stevens, W. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Tabachnick, B., & Fidell, L. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper Collins.

Taeschner, T. (1991). *A developmental psycholinguistic approach to second language teaching*. Norwood, NJ: Ablex.

Tahta, S., Wood, M., & Loewenthal, L. (1981). Age changes in the ability to replicate foreign pronunciation and intonation. *Language and Speech, 24,* 363–372.

Van de Guchte, C., Verhoeven, L., & Vermeer, A. (1986). *Taaltoets allochtone kinderen*. Tilburg: Zwijssen.

van Kleef, M. (1988). *Nederlandse taal voor kleuters in meertalige groepen: Knoop het in je oren*. Rotterdam: Projectbureau OVB Rotterdam.

van Kleef, M., & Boers, M. (1990). *Nederlandse taal voor kleuters in meertalige groepen: Laat wat van je horen*. Rotterdam: Projectbureau OVB Rotterdam.

Verhoeven, L., & Vermeer, A. (1992). Woordenschat van leerlingen in het basis- en MLK-onderwijs. *Pedagogische Studiën, 69,* 218–234.

Winitz, H. (1981). *The comprehension approach to foreign language teaching*. Rowley, MA: Newbury House.

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716401001011 Published online by Cambridge University Press