

J. Child Lang. 40 (2013), 567–585. © Cambridge University Press 2012
doi:10.1017/S0305000912000153

The use of the Internet in collecting CDI data – an example from Norway*

KRISTIAN E. KRISTOFFERSEN, HANNE GRAM
SIMONSEN

University of Oslo

DORTHE BLESES

University of Southern Denmark

SONJA WEHBERG

Odense University Hospital

RUNE NØRGÅRD JØRGENSEN

University of Southern Denmark

ELI ANNE EIESLAND AND LAILA YVONNE HENRIKSEN

University of Oslo

(Received 9 February 2011 – Revised 16 November 2011 – Accepted 24 March 2012 –
First published online 15 May 2012)

ABSTRACT

This article presents the methodology used in a population-based study of early communicative development in Norwegian children using an adaptation of the MacArthur-Bates communicative development inventories (CDI), comprising approximately 6500 children aged between 0;8 and 3;0. To our knowledge, this is the first CDI study collecting data via the Internet. After a short description of the procedures used in adapting the CDI to Norwegian and the selection of participants, we discuss the advantages and potential pitfalls of using

[*] We thank the CDI Advisory Board for permission to adapt the MacArthur-Bates Communicative Inventories to Norwegian. We also thank Master of linguistics Kristin Wium for drafting the first version of the Norwegian CDI, as well as PhD (linguistics) Janne von Koss Torkildsen, Lars Smith and Stephen von Tetzchner, professors of psychology at the University of Oslo, for evaluating the first draft. Furthermore, we would like to thank the parents of the participating children for having completed the CDI-reports. Finally, we thank the anonymous reviewers of the *Journal of Child Language* for their useful comments and suggestions. The data collection was funded by the Faculty of Humanities, University of Oslo. Address for correspondence: Kristian E. Kristoffersen, University of Oslo – Department of Linguistics and Scandinavian Studies, P.O. Box 1102 Blindern, Oslo N-0317, Norway. e-mail: k.e.kristoffersen@iln.uio.no

web-based forms as a method of data collection. We found that use of web-based forms was far less time-consuming, and therefore also far less expensive than the traditional paper-based forms. The risk of coding errors was virtually eliminated with this method. We conclude that in a society with high access to the Internet, this is a method well worth pursuing.

INTRODUCTION

Development of language skills in small children is characterized both by an enormous complexity – increasing with increasing age – and by extensive variation from one typically developing child to the next. How can we acquire knowledge of this development in a feasible way, identifying important milestones as well as covering the range of variation in the child population?

One way of addressing this question is by way of parental reports. Parents are the closest ones to observe their children's communicative skills, and it has therefore turned out that they are particularly reliable sources of knowledge about these skills. In addition, parent reports give information about linguistic skills across different situations, thus providing more representative data than can be obtained through structured tests or laboratory samples. Parent reports are also a cost efficient means for assessing linguistic skills in children, in particular for the early phases of development. Therefore they represent an invaluable tool for collecting the large samples that are necessary for establishing population-based norms, (see, e.g., Bates, Bretherton & Snyder, 1988; Berglund & Eriksson, 2000; Bleses *et al.*, 2008a; Dale, Bates, Reznick & Morriset, 1989; Fenson *et al.*, 1993; Fenson *et al.*, 1994; Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007; Meints, Plunkett & Harris, 1999; Reese & Reed, 2000).

One of the best-known and most used parent report instruments today focusing on development of gestures, vocabulary and grammar in infants and toddlers is the *MacArthur-Bates Communicative Development Inventories (CDI)* (Fenson *et al.*, 1994; Fenson *et al.*, 2007). The CDI was originally developed for children learning American English and has been adapted into a wide range of languages, spoken as well as signed; to date there exist more than fifty adaptations (Dale & Penfold, 2011). The CDI instrument has a long history in child language studies, see Fenson *et al.* (2007: 47–50) for an overview of the development and different versions of the American original, and Law and Roy (2008) for a recent review of research based on CDI reports within various languages. These adaptations differ from the American original in a number of ways, reflecting cultural and linguistic differences. However, the fact that they have been adapted from the same original makes them a good starting point for cross-linguistic

studies (Bleses *et al.*, 2008b; Caselli *et al.*, 1995; Caselli, Casadio & Bates 1999; Caselli, Monaco, Trasciani & Vicari, 2008; Devescovi, Caselli, Marchione, Pasqualetti, Reilly & Bates, 2005; Maital, Dromi, Sagi & Bornstein, 2000; McBride-Chang *et al.*, 2008; Tardif, Gelman & Xu, 1999). The CDI forms have also been used in investigations of language skills in children from atypical populations (see Law & Roy (2008) for a review).

In its present version the MacArthur-Bates CDI consists of two different forms, an infant form (*Words and Gestures*) covering development between 0;8 and 1;4, and a toddler form (*Words and Sentences*) covering the period from 1;4 to 2;6. Each form has a number of different sections, covering a range of communicative skills. In the infant form there are two main parts, 'Early words' and 'Actions and gestures'. In 'Early words', first signs of understanding, productive skills like labelling and imitation, and size of receptive and productive vocabulary are assessed. The second part, as the name implies, focuses on communicative actions and gestures.

The toddler form also has two main parts. In the first part, 'Words children use', there is an extensive vocabulary checklist assessing productive vocabulary in addition to a section focusing on the way children use words to talk about past and future actions, as well as absent objects and persons. The second part ('Sentences and grammar') focuses on inflections, overgeneralizations and grammatical complexity.

Recently, we made an adaptation of the MacArthur-Bates CDI into Norwegian, and used it in a large-scale population-based study of children aged 0;8 to 3;0 learning Norwegian. The most innovative aspect of this study was the method of data collection: the data were collected on the Internet. In this article we will address methodological aspects of collecting data in this way.

Web-based data collection

Traditionally, data collection in CDI-based studies has been paper-based, i.e. parents have completed the reports on paper. In most cases the report forms have been scored, coded and entered into a database manually, as was the case with, e.g., the Danish CDI study (Bleses *et al.*, 2008a), or they have been scanned and scored automatically, as was the case with, e.g., the US study (Fenson *et al.*, 2007). Today, when a growing number of people have access to the Internet, new possibilities for collection of research data open up. Therefore, we decided to collect our data for this study via the Internet. In this way, we would be able to explore the possibilities of this methodology, and compare it to the more traditional means of collecting data. To our knowledge this is the first time CDI data has been collected via the Internet.

Evident advantages of web-based data collection are that it is cost-efficient and speedy. Also, coding errors are virtually eliminated. One potential challenge is that only certain groups of the population have access to the Internet, resulting in an unrepresentative composition of the sample. Accordingly, web-based data collection can only be successful in societies where a comparatively large rate of the population has access to the Internet. In the Nordic countries rate of access is high, and according to Eurostat (2008) 84 percent of the Norwegian population had access to the Internet in 2008. This is a high proportion as compared with an average of 60 percent in all European countries. Still, high access to the Internet does not necessarily mean that a study using web-based data collection will result in the same response rate as would a study using data collection by paper copies of the forms. Fortunately, we are in a very good position to compare response rates from our own web-based study with a similar study using a more traditional methodology in a society comparable to Norway's, Denmark (Bleses *et al.*, 2008a).

Another challenge arising from collecting data via the Internet is that the composition of the sample may be skewed in the direction of higher parental education. An over-representation of children with parents with higher education has been observed in other CDI studies, including the US and the Danish ones (Bleses *et al.*, 2008a; Fenson *et al.*, 2007). Thus, one may ask to what extent web-based data collection adds to this skewness towards higher education.

Against this background, we address the following research questions: (1) How does the response rate of the present study using web-based data collection compare to a CDI study using a more traditional way of data collection (Danish); (2) To what extent can web-based data collection lead to an unrepresentative composition of the sample in favour of higher parental education; and (3) To what extent did the respondents experience problems of various sorts directly connected to the web-based design?

Before approaching these questions we will briefly describe how the instrument was adapted into Norwegian, how the validity of the items was examined, and how the participants were selected.

THE NORWEGIAN ADAPTATION OF THE CDI

Norwegian is a Germanic language spoken by ca. 4,985,000 people (estimated population by 1 January 2012; Statistics Norway, 2011), with a range of different dialects. Most closely related to Norwegian are the other Scandinavian languages Swedish and Danish, more distant relatives are Icelandic and Faroese. The Norwegian lexicon is predominantly Germanic in origin, with loan words coming from a range of different source languages. Morphologically, Norwegian is slightly more complex than

English. Nouns fall in two or three (depending on dialects) gender classes, and are inflected for number and definiteness. Like verbs in the other Germanic languages Norwegian verbs fall in two main classes, weak ('regular') and strong ('irregular'). However, there are two weak classes in Norwegian (and in the other Scandinavian languages), as opposed to only one in English. Norwegian verbs are inflected for tense, mood and voice. Adjectives are inflected for number, definiteness and gender. The main mode of inflection is suffixation.

Work with the Norwegian adaptation of the CDI started in 2006, when a first version was constructed on the basis of a comparison with the American original, and evaluated by a group of experts on early communicative development in Norwegian children within the fields of linguistics and psychology. In order to evaluate the appropriateness of the inventory items selected as well as the instructions given to the parents, a pilot study was conducted in 2007. Parental report data from seventeen children were collected, six with the *Words and Gestures* form, and eleven with the *Words and Sentences* form. In addition to completing the CDI forms the parents participating in the pilot study provided information on family relations, the child's contact with other languages and her/his medical history. Also, the parents reported the time spent on completing the forms and evaluated the instructions given in the forms. Finally, the participating parents were asked to add vocabulary items that they felt were missing from the forms.

All the parents who participated in the pilot study reported that they found the instructions clear and easy to understand. They required between 10 and 80 minutes to complete the forms. A few words were added (for example *pc*) to the vocabulary sections in response to suggestions from some of the parents, and a few others were removed. In all, in the second version the Norwegian CDI forms were only slightly revised.

As a final step, however, before constructing the web-based forms used in the present study we revised the forms again, this time aiming to bring the Norwegian adaptation as close as possible to the Danish adaptation, in order to facilitate cross-linguistic comparison between two closely related languages with comparable grammatical systems but with quite different phonologies. A comparison between the vocabulary sections of the third version of the Norwegian adaptation of the CDI, the American English original, and the Danish adaptation can be found in Table 1.

There are also differences between the Norwegian, Danish and American CDIs concerning the items focusing on grammatical skills. In the section 'Word endings, part 1', the Norwegian CDI has six items, in comparison with four items in the American and three items in the Danish. The difference between the Danish and American CDIs is due to one extra item in the American CDI for the present progressive, a form that does not exist in either of the two Scandinavian languages. The additional two items in the

TABLE 1. *Comparison of the categories and number of items in the vocabulary lists of the Norwegian, Danish and American CDIs*

	CDI: Words and gestures			CDI: Words and sentences		
	<i>Norwegian</i>	<i>American</i>	<i>Danish</i>	<i>Norwegian</i>	<i>American</i>	<i>Danish</i>
Total vocabulary	396	396	410	731	680	725
1 Sound effects and animal sounds	11	12	11	12	12	12
2 Animals (real or toy)	36	36	36	44	43	43
3 Vehicles (real or toy)	9	9	10	14	14	14
4 Toys	8	8	8	18	18	18
5 Food and drink	28	30	28	68	68	68
6 Clothing	20	19	21	30	28	30
7 Body parts	20	20	20	27	27	28
8 Small household items	35	36	39	50	50	50
9 Furniture and rooms	23	24	24	34	33	33
10 Outside things	14	27	14	31	31	31
11 Places to go	12	—	14	22	22	22
12 People	22	20	30	36	29	40
13 Games and routines	18	19	15	27	25	27
14 Action words	53	55	53	108	103	103
15 Words about time	8	8	8	16	12	15
16 Descriptive words	36	37	36	62	63	63
17 Pronouns	11	11	11	31	25	31
18 Question words	6	6	6	7	7	7
19 Prepositions and locations	16	11	16	41	26	41
20 Quantifiers and articles	9	8	10	22	17	21
21 Auxiliary verbs	—	—	—	22	21	21
22 Connecting words	—	—	—	9	7	6

Norwegian version are concerned with past participle and definiteness forms. The motivation for including definiteness is that, unlike English, this grammatical category is expressed inflectionally in Danish and Norwegian. Also the section 'Word forms (nouns and verbs)' differs in the three versions: the American original and the Norwegian adaptation have irregular nouns and verbs only, whereas Danish also includes regular nouns. Finally, the section 'Word endings, part 2 (nouns and verbs)' differs in the three versions in that the forms for the two Scandinavian languages have more items than the US original. The main reason for this difference is that both nouns and verbs have more than one inflectional class in Danish and Norwegian, resulting in more inflectional variation.

VALIDITY OF THE ITEMS IN THE NORWEGIAN CDI

Several studies have examined the validity of the CDI instrument and have found it acceptable (see, e.g., Berglund & Eriksson, 2000; Bleses *et al.*,

2008a; Fenson *et al.*, 1994; Reese & Reed, 2000; Thordardottir & Ellis Weismer, 1996). For example, Bleses *et al.* (2008a; Bleses, Vach, Wehberg, Faber & Madsen, 2007) investigated the validity of the Danish CDI-instrument (i) by comparing words spontaneously produced by Danish children and words in the vocabulary list in *CDI: Words and Sentences*, (ii) by comparing words found in the three longest sentences produced by each child and the words in the vocabulary list in *CDI: Words and Sentences*, and (iii) by correlating vocabulary size measured in spontaneous speech samples with vocabulary size measured by CDI parental reports. For all analyses, they found acceptable validity – 73 percent of relatively frequent words (found at least five times) in the spontaneous speech corpus and 75 percent in the three longest sentences were also included in the CDI list; and there was a high correlation between the CDI scores and vocabulary size in spontaneous speech for four children (Bleses *et al.*, 2008a: 657f.).

To check the validity of the selection of the vocabulary items in the Norwegian CDI two investigations were made, corresponding to (i) and (ii) above. In the first investigation we compared these items with the lexical items included in a longitudinal corpus of Norwegian child language, originally collected by Simonsen (1990) for phonological analysis. This corpus contains spontaneous speech data from three children, two girls and one boy, covering the age span from two to four years. All three children have grown up in Oslo and speak the variety of Norwegian known as Urban East Norwegian. Data were recorded as a dialogue between the researcher and the child, in play situations with toys and books, sometimes with the mother present, sometimes not. Since the main purpose of that investigation was to explore the phonological development of the children, the play material was chosen to facilitate elicitation of examples of all Norwegian speech sounds in all positions, including pictures and toys of relatively infrequent words. This may have skewed the vocabulary somewhat in the direction of particular semantic categories. Some of the categories covered in the CDI, for example those related to food and drink, were not naturally covered in the play situations in the recordings. Furthermore, the recordings were made in the mid 1980s, when words like, e.g., *pc* and *trampoline* were less common.

Each session lasted between 30 minutes and 1.5 hours. To be able to compare the children with respect to vocabulary frequency, samples of approximately 150 different word forms (types) produced by the children – varying between 121 and 180 types – from each session were extracted (excluding names and direct imitations). For the present comparison, only the data points below three years were included, (11 samples in all from 2;0 to 3;0). The samples include a total of 3,997 word tokens produced by the children, distributed on 651 different word types. Because of the age range, we have only compared them with the vocabulary listed in *Words and Sentences*.

Comparing these word types to the word types included in the Norwegian CDI (*Words and Sentences*), we found that 58 percent of these words were also found in the CDI. When we removed all words that the children had produced only once, the percentage rose to 72.6 percent, and reducing the number of words to those produced at least five times, the percentage rose even higher, to 78.6 percent. Thus, more than three quarters of the vocabulary items used frequently by these children were found in the CDI – a percentage matching the Danish results (Bleses *et al.*, 2008a). The fifty most frequent words found in the Simonsen dataset had – not unexpectedly – a clear over-representation of function words as compared to content words. Only eighteen of the words were content words: two nouns, three adjectives and thirteen verbs, among which six were auxiliaries. The remaining thirty-two words were function words, among which thirteen were pronouns, nine adverbs, and the remaining words evenly distributed on prepositions, conjunctions, interjections and determiners. Only four of these fifty words were not represented in the CDI: the three function words *sånn* ‘like this’, *da* ‘then’, *bare* ‘only’, and the verb *komme* ‘come’. In retrospect, the frequent verb *komme* ‘come’ might have been included in the CDI – on the other hand, it was not included in, e.g., the Danish or American version, so for comparative purposes it was defensible not to include it.

The second investigation of the validity of the Norwegian CDI involved a comparison between the vocabulary items in the CDI and the vocabulary in the sentences reported by the parents in their response to the question about the three longest sentences the child had produced (cf. (ii) above).

We chose four different time points – 1;6 ($N=182$), 1;11 ($N=211$), 2;1 ($N=227$), and 2;4 ($N=187$) – and for each time point we extracted all the words in the three longest sentences reported for the children. We grouped the word types according to frequency (all words, words produced at least twice, words produced at least five times) and compared them with the vocabulary items in the *CDI: Words and Sentences*. The results show that for the youngest age group (1;6), 73% of all words produced were also found in the CDI, but for the older age groups, with more words produced, the percentage of words also found in the CDI decreased to between 52% and 66%. However, the more frequently the words were produced, the higher the overlap with the vocabulary found in the CDI. Already for words produced at least twice, the percentage of words found also in the CDI was never below 78%, and for words produced at least five times, the percentage found also in the CDI varied between 85% and 93%. Finally, for words produced at least eighteen times, the percentage of words found also in the CDI varied between 100% for the youngest children and 94% for the oldest ones. This last count was made to compare the results

from the Norwegian CDI with those from the Swedish and Danish CDIs (Berglund & Eriksson, 2000; Bleses *et al.*, 2008a). The rates of overlap between frequently produced words and the CDI vocabulary lists in the three languages were comparable – slightly higher for the Danish and Norwegian data than for the Swedish ones (see Appendix A). The words missing among these frequently produced words for Norwegian were the verb *komme* ‘come’, the noun *barnehage* ‘kindergarten’ and the infinitive marker *å* ‘to’.

METHODS

Selection of participants

In October 2008 the web version of the parental forms was constructed by the Danish company MikroVærkstedet (in collaboration with Center for Child Language at the University of Southern Denmark). During the latter half of that month the first version of the web forms was tested by several members of the project staff.

In November 2008 Statistics Norway (the official Norwegian statistical agency) randomly selected 20,400 families with children aged between 0;8 and 3;0 from the official Norwegian birth register. Since information about individuals would be handled in the study, the Norwegian Social Science Data Service (NSD) reviewed the methods for collecting and storing data and eventually approved all procedures. The procedures were also evaluated and found appropriate by Statistics Norway. The children had to be Norwegian citizens and have the exact age of 0;8, 0;9, 0;10 ... or 3;0 between the dates 21 November and 28 November 2008. All the selected families received a letter describing the study and inviting them to participate. The letter was sent through Statistics Norway, who took care of the anonymity of the parents. The letter also provided detailed instructions for accessing the web-based forms, as well as an individual user name and password.

For those who decided to take part in the study more information was given on the website. Among other things, the parents were asked to indicate (1) whether they wanted to participate in a longitudinal study by sending in monthly reports, (2) whether they would like to participate in a lottery with the possibility to win a gift certificate, and (3) whether they wanted to receive a profile of their child’s linguistic skills at the appropriate age level.

By 1 January 2009, 5,315 forms were completed. Then, in the third week of January 2009 Statistics Norway sent a reminder by regular mail to the more than 14,500 families who had not completed the forms in the first round, either because they never started, or because they were prevented from completing them for technical or other reasons.

TABLE 2. *Gender and sibling status of the participants, as compared to the general population*

	Norwegian CDI		Child family population in Norway ^a	General population in Norway
Gender	N	%	%	%
Boys	3225	49	51	50
Girls	3349	51	49	50
Sibling status				
Only child	2226	34	37	18 ^b
At least one sibling	4348	66	63	82

^a Based on the age groups 0, 1, 2 years.

^b Among children living at home, 0–17 years.

By 8 March 2009, i.e. after the second wave of data, a total of 7,555 forms had been completed, with 2,699 for *Words and Gestures* and 4,856 for *Words and Sentences*, yielding a response rate of 37%. Beforehand Statistics Norway had estimated a response rate somewhere between 35 and 50%. The response rate varied between each monthly stage, with 22% as an extreme at the lowest end for the children aged 0;8, and 54% as an extreme at the highest end for the children aged 1;9. Generally, the response rate seemed to increase with the age of the child. (See Appendix B for an overview of responses at each monthly age.)

Table 2 shows the gender and sibling status of the participants in the study, compared to those of the child family population in Norway and the general population in Norway (all information obtained from Statistics Norway). The sample is balanced with respect to gender (49% boys and 51% girls), and the sibling status of the children matches that of the child family population relatively well.

As already mentioned, Norwegian has a wide dialect variation. The dialects can be grouped into four main categories: East Norwegian (including the capital Oslo), South and West Norwegian, Trøndelag Norwegian and North Norwegian. Participating families came from all these dialect areas, in a proportion that corresponded very closely (<1% difference) to the population in these areas.

Web-based data collection

The project was administered from a website where both the web-forms and information about the project were available. The amount of information was kept at a level that reflected both expected information needs and reader usability, in the sense that the amount of text was kept to a minimum. The website was programmed in PHP, HTML and CSS, using

the Zend Framework. All data was stored in a MySQL-database on a standard Linux server. At the time when the web-forms were made available for the participants, two research assistants were ready to answer questions from the participants. The research assistants could be contacted through e-mail or telephone. All queries were answered as soon as possible, in most cases within the next 24 hours.

While information about the project was openly available at the website, only participants could access the actual forms with a username and a password. All potential participants were created as users in a database on the basis of data from Statistics Norway. All in all 20,400 users were created with birthday of the child (username) and a serial number (password). Upon entering the website the parents would on the basis of their username automatically be directed to the right form. Accordingly, all children between 0;8 and 1;4 were directed to the *Words and Gestures* form, whereas the children between 1;8 and 3;0 were directed to the *Word and Sentences* form. The children between 1;4 and 1;8 were randomly directed to either the *Words and Gestures* or the *Words and Sentences* form, resulting in two groups of equal size.

In order to complete data submission the participating parents had to go through three successive steps. The first was to indicate the gender of the child and the status relative to the child of the person who was completing the CDI-forms (mother, father or both). The second step was to complete the relevant CDI-form, and the third and final step was to complete a background information questionnaire addressing a number of socio-demographic factors: the child's (present and earlier) place of residence, sibling status, contact with other languages, birth and health information, and information about the parents (place of residence, age, level of education).

At any time during the session the parents could log out of the system and then return and finish later on. In that case they would be directed to the place in the form where they broke off. Furthermore, the user was allowed to move back and forth in the form and correct responses until pressing the final submit button. However, this did only work within the CDI-form – once the user had reached step three (the background information questionnaire), it was no longer possible to return to step two (see below).

The exact age for the child was calculated from the moment the person completing the forms finished step two and entered step three.

Exclusion criteria and final dataset

A set of criteria was applied to ensure that the children included in the final dataset were monolingual Norwegian speakers without any known health problems. To be included in the final dataset, then, the child had to satisfy

the following criteria: (1) no frequent contact with other languages than Norwegian; (2) birth at full term (after week 36); (3) combined hospital stay should not exceed 4 weeks; and (4) no serious, well-founded parental concern for the language development of the child. This meant that children with limited hearing because of frequent hearing infections were not excluded, but profoundly deaf children were, as well as children who had physical or mental disadvantages in learning to speak, or cases where daycare personnel also had raised concern. No more than forty children were excluded by this fourth criterion.

Furthermore, the age of the child had to be between 0;8 and 1;8 for the infants, and between 1;4 and 3;0 for the toddlers, and at least one question in the form had to be answered. After applying these criteria, 981 of the original 7,555 children were excluded, so the final dataset consisted of 6,574 parental reports, 2,359 for *Words and Gestures*, and 4,215 for *Words and Sentences*.

RESULTS AND DISCUSSION

Response rates

As for response rates, at least as far as the Norwegian and Danish CDIs are concerned, the medium of the forms – web-based or paper-based – did not seem to matter. Response rates for the two studies were quite similar: 37% for the Norwegian CDI study and 34% for the Danish cross-sectional CDI study (Bleses *et al.*, 2008a: 655).

In our study, the sample was also skewed in the direction of higher educational levels of the parents responding. However, when compared to the Danish CDI, which was paper-based, the results are comparable – and when we take the child family populations in the respective countries into account, there is actually a better match in the Norwegian study than in the Danish study (see Table 3).

Problems experienced by parents with the web-based design

We were interested in knowing to what extent users experienced problems related to the web-based design. As mentioned above, the participants could get in contact with research assistants as they were working with the forms and report their problems. In addition to the important – and primary – effect of assisting the participants, this made it possible for us to identify the kind of problems they experienced.

In the time period from November 2008 to February 2009 we received 157 queries. Of these 21 were from individuals not participating in the study. Among these were journalists, speech therapists and students who wanted more information, in addition to members of the general

TABLE 3. *Educational levels of parents responding in the Danish and Norwegian CDI studies^a*

	Norwegian CDI		Child family population in Norway	Danish CDI		Child family population in Denmark
	N	%	%	N	%	%
Parent education						
Basic (<9 years)	453	3	11	94	4	28
Short further education (12 years)	3949	30	32	1005	42	52
Medium further education (14–16 years)	5524	42	37	680	28	11
Long further education (> 16 years)	3134	24	16	499	21	9

^a Note that the number of responding parents is larger in the Norwegian than in the Danish study. The reason for this is that in the Norwegian study information about educational level was available for both parents of all participants, whereas in the Danish study this information was available for only a subset of the parents.

public who wanted to participate with their children. The remaining 136 questions were from participants in the study. We will focus on the latter type of queries in the following. More specifically, these were the questions that were caused by a lack of information in the letter or on the website, or by the architecture of the website. After the first wave of data collection, the most common questions were used to make a list of frequently asked questions (FAQs), which was posted on the website.

The queries can be divided into three general categories: ‘Technical issues’ (46 queries), ‘Requests for practical information’ (73 queries), and ‘Comments about the website’ (17 queries).

Technical issues

The forty-six queries in the category ‘Technical issues’ were mostly from participants who experienced problems with the website itself. Some users reported not being able to access the site or log in. As far as we could tell this was due to user error.

The largest number of queries in this category (16 queries) came from users who wanted to re-enter and change their form. Many of these participants were having problems because the form allowed the user to go back to previous pages, but only within the CDI form. Once the respondents had started completing the background information questionnaire, it was no longer possible to go back. These participants were asked to wait until January, when the second round would start, and then fill out the form again.

Another common query came from participants who had forgotten to fill in their e-mail address, or to tick the boxes that indicated that they wanted to participate again, or to receive a 'linguistic profile' of their child. These participants could have logged back in and fixed the problem on their own, but since it demanded little time and effort we did this manually for them.

Practical information

As for the seventy-three queries in the category 'Practical information', most of these were questions about how to fill out the form and how to log in. Some participants had lost their login information, and two wanted a paper version of the form. There were also some questions about the 'linguistic profile' that parents could choose to receive for their child. Many parents expected to receive it soon or immediately after completing the form. In reality, the profiles could not be made available until a few months later, when all the data had been collected and processed. This was not made clear on the website, and had to be explained to those who were waiting for it to come.

We received sixteen queries from parents with twins or triplets. Since the invitation letter from Statistics Norway only included a birth date, and not the name of the child, they were unsure about which child they were meant to complete the form for. They were instructed to pick one child randomly.

Comments

The third category of queries (17) were comments from the participants about various aspects of the study – either about the technical solution, the contents of the form, the way the form was worded or the way the letters were sent out. Some participants had suggestions for words they felt should be included in the form. Others commented on dialect differences in vocabulary between the CDI items and their child's own words. Several of these commented that the differences made the form difficult to fill out, or that it gave a wrong picture of their child's vocabulary.

Another common concern came from parents who felt that the form was too extensive for their child. Many of these parents had a child in the lower age group for the form they had received (0;8–0;9 for the infant form, 1;4–1;8 for the toddler form), and were frustrated by hardly being able to fill in anything at all. We answered these queries in some detail, both to avoid unnecessary concerns about their child's language development, and to persuade those who had not finished the forms to do so.

CONCLUSIONS

As for the adaptation to Norwegian in general, we assume that it is close enough to both the US and the Danish versions to make a good basis for cross-linguistic comparisons. The validity investigations also indicate that the choice of vocabulary items in the Norwegian CDI is acceptable, and comparable to the results for the Danish and Swedish CDIs. However, the most interesting methodological aspects are related to the web-based design. What are the advantages with using the Internet for data collection, and what are the limitations?

This method represented an extremely efficient way of collecting a large amount of data, in terms of both speed and accuracy of coding. In this Norwegian study, the time spent between sending out the invitation letters to the participants and the calculation of the final norms was approximately four months (mid-November 2008 to mid-March 2009). In comparison, the Danish, paper-based cross-sectional CDI study, which was completed in 2006, took more than four years to obtain the results (April 2002 – September 2006) (Bleses *et al.*, 2008a). In the Danish study, ten research assistants spent more than two years to enter, code and control the data. In the Norwegian study, there were no intervening levels between parental reporting and entry into the database, so there was no need for assistants for this kind of work, and no risk of coding errors.

Our concern that data collection via the Internet would result in a lower response rate and a more skewed composition of the sample of respondents in the direction of higher parental education turned out not to be confirmed. We found that the response rate was about the same as in the Danish CDI study, comprising a comparable number of participants, and that the skewness among participants towards higher education in the parents was no higher than in the Danish study (actually slightly less skewed). Geographically, there did not seem to be any skewness, either. However, it is important to remember that such results can only be obtained in countries with high access to the Internet – thus for the time being the use of web-based questionnaires such as these should be limited to such countries.

We had two assistants ready to answer questions by e-mail and phone during the period of data collection. This way, we were able to provide quick feedback to the participants to problems they experienced when completing the forms. This seems to be a good idea for several reasons. Although the rate of participants reporting problems was low – with 7,555 completed forms, 136 queries yield a surprisingly low percentage (1.8%) of parents experiencing problems – this possibility may have reduced the risk of parents giving up for technical or other reasons. Having the possibility of direct contact with the parents is in itself a clear improvement from a paper-based survey in a letter. It also gave us a chance to hear parents'

reactions on all parts of the study – reactions to be taken into account when evaluating the results, and as reminders for improvements in further use of the instrument.

Only one-third of these queries were related to the web format, indicating that the web format in itself was not a serious obstacle for the parents to take part in the study. However, missing information about the study, in particular as to how and when the parents would obtain a linguistic profile of their child, clearly is something to improve. In retrospect, information about the estimated point of time when the profiles would be ready – and only after the whole study had been finished – should have been included in the letter or on the website. The fact that it is possible to get an indication of the child's performance relative to his or her age-mates is an interesting side effect of our web-based design. However, it is important to remind the parents of the huge variation and changes in performance with age.

Quite a few comments were related to the choice of vocabulary items, and in particular to dialect differences that made the forms difficult to fill out for some parents. As indicated above, participants came from all dialect areas in Norway, in approximately the same proportion as in the general population. Our answers to those parents were in line with the general instructions for CDIs – we understand that in some cases the parents felt that these items do not give a correct picture of their child's real competence, but across dialects this will probably even out.

Parents giving up because their children are in the youngest segment and therefore have only few items to mark is a potential source of error. As indicated above, the response rate tended to be lower among the younger children than among the older ones. We stressed the importance of investigating the full range of variation to these parents but have not found a good way of solving this problem – however, again this is probably not related to the web design per se.

In sum, we conclude that the advantages outweigh the possible problems, so in countries with high access to the Internet in the population, this method of data collection is worth pursuing.

REFERENCES

- Bates, E., Bretherton, I. & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.
- Berglund, E. & Eriksson, M. (2000). Communicative development in Swedish children 16–28 months old: The Swedish early communicative development inventory: Words and Sentences. *Scandinavian Journal of Psychology* **41**(2), 133–44.
- Bleses, D., Vach, W., Wehberg, S., Faber, K. & Madsen, T. O. (2007). *Tidlig kommunikativ udvikling*. Odense: Syddansk universitetsforlag.
- Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O. & Basbøll, H. (2008a). The Danish Communicative Developmental Inventories: Validity and main developmental trends. *Journal of Child Language* **35**, 651–69.

- Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O. & Basbøll, H. (2008b). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language* **35**, 619–50.
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L. & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development* **10**, 159–200.
- Caselli, M. C., Casadio, P. & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language* **26**, 69–111.
- Caselli, M. C., Monaco, L., Trasciani, M. & Vicari, S. (2008). Language in Italian children with Down syndrome and with specific language impairment. *Neuropsychology* **22**, 27–35.
- Dale, P. S., Bates, E., Reznick, J. S. & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language* **16**, 239–51.
- Dale, P. S. & Penfold, M. (2011). Adaptations of the MacArthur-Bates CDI into non-U.S. English languages. 2011(1114). Retrieved from www.sci.sdsu.edu/cdi/documents/AdaptationsSurvey7-5-11Web.pdf.
- Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J. & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language* **32**, 759–86.
- Eurostat. (2008). Internet access and use in the EU27 in 2008. Retrieved 26 August 2010, from <http://europa.eu/rapid/pressReleasesAction.do?reference=STAT/08/169&format=HTML&aged=0&language=EN&guiLanguage=nl>.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B. & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59**(5). Chicago: University of Chicago Press.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego: Singular Publishing Group.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S. & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories. User's guide and technical manual*. Baltimore: Brookes.
- Law, J. & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child and Adolescent Mental Health* **13**, 198–206.
- Maital, S. L., Dromi, E., Sagi, A. & Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language* **27**, 43–67.
- McBride-Chang, C., Tardif, T., Cho, J.-R., Shu, H., Fletcher, P., Stokes, S. F., Wong, A. & Leung, K. (2008). What's in a word? Morphological awareness and vocabulary knowledge in three languages. *Applied psycholinguistics* **29**, 437–62.
- Meints, K., Plunkett, K. & Harris, P. L. (1999). When does an ostrich become a bird: The role of prototypes in early word comprehension. *Developmental Psychology* **35**, 1072–78.
- Reese, E. & Reed, S. (2000). Predictive validity of the New Zealand MacArthur communicative development inventory: Words and sentences. *Journal of Child Language* **27**, 255–66.
- Simonsen, H. G. (1990). Barns fonologi: system og variasjon hos tre norske og et samoisk barn. Unpublished doctoral dissertation, University of Oslo.
- Statistics Norway (2011). Population Statistics. Estimated population, 1 January 2012. Retrieved from www.ssb.no/english/subjects/02/01/10/folkber_en/.
- Tardif, T., Gelman, S. A. & Xu, F. (1999). Putting the 'noun bias' in context: A comparison of English and Mandarin. *Child Development* **70**, 620–35.
- Thordardottir, E. T. & Ellis Weismer, S. (1996). Language assessment via parent report: Development of a screening instrument for Icelandic children. *First Language* **16**, 265–85.

APPENDIX A

Words in MLU₃ at four time points compared to vocabulary items in the Norwegian, Danish and Swedish CDI: Words and sentences

		Word types, frequency across children			
		1;6/1;7 ^b	1;11	2;1	2;4
Norwegian CDI	Words produced at least 18 times	11	38	39	49
	Words included in the CDI (%)	100%	97%	95%	94%
Danish CDI	Words produced at least 18 times	18	29	33	38
	Words included in the CDI (%)	100%	97%	97%	97%
Swedish CDI	Words produced at least 6 times	12	26	39	55
	Words included in the CDI (%)	83%	92%	97%	87%

^a While the Norwegian and Danish datasets are comparable in size, the Swedish dataset is more than three times smaller, which means that a token frequency of six in the Swedish dataset is roughly equivalent to eighteen in the Norwegian and Danish datasets.

^b The Norwegian data are from children aged 1;6, while the Danish and Swedish data are from children aged 1;7.

APPENDIX B

Potential and actual participants at each monthly stage

Age	Potential participants	All completed forms	Forms included in the final dataset	Response rate – all completed forms	Response rate – forms included in the final dataset
8	610	138	124	22%	20%
9	582	185	159	32%	27%
10	601	226	202	38%	34%
11	601	192	165	32%	27%
12	601	207	185	34%	31%
13	601	225	194	37%	32%
14	601	213	186	35%	31%
15	590	217	191	37%	32%
16	1172	362	308	31%	26%
17	1205	393	345	33%	29%
18	1208	440	377	36%	31%
19	1175	457	400	39%	34%
20	1256	461	403	37%	32%

THE USE OF THE INTERNET IN COLLECTING CDI DATA

21	601	326	270	54%	45%
22	601	293	260	48%	43%
23	601	247	211	41%	35%
24	601	221	195	37%	32%
25	601	254	227	42%	38%
26	601	245	207	41%	34%
27	587	210	188	36%	32%
28	595	232	187	39%	31%
29	602	230	196	38%	33%
30	596	244	216	41%	36%
31	609	235	210	39%	34%
32	607	220	193	36%	31%
33	601	240	211	40%	35%
34	601	212	183	35%	30%
35	601	238	211	40%	35%
36	601	192	170	32%	28%
	20,409 ^a	7,555	6,574		

^a This number includes nine fictional test children whose data were entered in order to test the web-based forms before data collection started. These data were later excluded.