German

DIALANG

Dansk
Deutsch
Ελληνικα
English
Español
Suomi
Français
Gaeilge
Islenska
Italiano
Nederlands
Norsk
Português
Svenska

INDEX

Introduction to DIALANG 2

Information about the Dialang Items 6

Dialang German Reading Comprehension Items 8

Dialang German Listening Comprehension Items 15

Introduction to DIALANG

Context

DIALANG is a computer-based and computer-scored suite of diagnostic tests of foreign language abilities in 14 European languages. There are tests of Reading, Writing, Listening, Structures and Vocabulary, in Danish, Dutch, English, Finnish, French, German, Greek, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish and Swedish. In addition, there is a Vocabulary Size Placement Test for each language, and, also for each language, there are self-assessment 'I Can' statements for Reading, Writing and Listening. DIALANG provides extensive feedback to users, and advice on how they might improve their abilities.

The DIALANG Project was funded by the European Commission under the Socrates LINGUA programme, ACTION D. The initial development of DIALANG, beginning at the end of 1996, was the result of collaboration among over 20 partner institutions, most of them universities, throughout Europe, coordinated by the University of Jyväskylä. For full details, visit the DIALANG website at www.dialang.org. The piloting, calibration and further development of the system was coordinated by the Free University of Berlin. All tests can be downloaded from the DIALANG website, free of charge, and can be taken over the Internet, again free of charge, from any computer on which the program is installed.

DIALANG is intended to help learners diagnose their strengths and weaknesses in any or all of the five abilities tested, to encourage learners to self-assess their abilities, and to compare their self-assessment with their test results. By so doing it is intended to enhance learner autonomy by encouraging learner reflection on what they know and can do in their target foreign language, and thereby to increase their awareness of what is involved in learning or improving in a foreign language.

The DIALANG Project began in 1996, when the first draft of the Council of Europe's Common European Framework of Reference (CEFR) was already available, and so it was possible to base the DIALANG Assessment Framework and Specifications on the CEFR. DIALANG also used the 'Can Do' statements contained in the CEFR, but adjusted their wording to suit its self-assessment purposes, essentially by changing 'Can-Do' into 'I Can' and at times simplifying the language to make it more accessible to learners. In addition, extensive explanations of the feedback given by the DIALANG system were based on the CEFR scales, and the advice provided to learners on how to improve their ability to use the language used the information contained in the CEFR's Descriptive Scheme (which became Chapters 4, 5, 6, and 7 in the 2001 publication - Council of Europe, 2001).

DIALANG not only contains tests in its 14 languages, but it also enables users to read test instructions and rubrics, help facilities, introductions and, most importantly, all self-assessment statements, feedback and advice, in any of these 14 languages. Once agreement had been reached on the appropriate wording, particularly of the 'I Can' statements, and the Explanatory and Advisory Feedback, these were then translated from the original English into the 13 languages of DIALANG, and checked for the quality of the translations. Thus DIALANG is multilingual in both the tests available, and in the interface and administration of the system.

Although intended to offer detailed diagnosis of language skills and subskills (for full details of these, see Alderson, 2005), DIALANG also reports the learner's ability in the skill or aspect of language being tested, as a level on the CEFR (A1 to C2), rather than as a numerical score.

DIALANG is intended to contribute to life-long learning and learner autonomy, and thus is "aimed" at any adult learner of foreign languages, whether or not that learner is engaged in institution-based language learning. However, many institutions consider the DIALANG system to be of value in placing students into language classes which are based on the CEFR, and thus, although intended to be diagnostic, the tests are increasingly being used for placement purposes. At the time of writing, roughly 1,500 users take a DIALANG test every day.

Test development process

a) Content quality

The DIALANG tests were developed by Assessment Development Teams (ADTs) for each of the 14 languages. Typically a team was composed of language educationalists, usually at institutions of adult or higher education. Members were experienced teachers, researchers and assessment experts, where these existed for the language in question. In several cases, where the traditions of assessment were different from those espoused in DIALANG and the CEFR, familiarisation with the CEFR and with appropriate methods of test and item writing was made available and a number of training sessions were held. The tests themselves, as already noted, were based on the DIALANG Assessment Framework (DAF) and DIALANG Assessment Specifications (DAS), themselves based on or derived from the CEFR and related User Guides. Detailed Guidelines for Item Writers were also developed and made available as the formats (test methods) of the tests and an associated computer-based authoring system became available.

Test development was coordinated by two Test Development Coordinators, who were responsible for the issue and updating of all documentation and guidance, and who communicated with the ADTs through the ADT Team Leader. As items were developed, first on paper and then over the Internet using the Authoring software developed by the Project, they were subject to an extensive process of quality review, first by peers within the ADT, then by the Team Leader, and then by external reviewers commissioned by the Test Development Coordinators, using the Internet-based Item Review software. Once revisions had been made and reviewed by the ADT, the Test Development Coordinators made a second review to ensure that all items conformed to the DAF and DAS, and to the requirements of the computer programs. In this way, large pools of items were developed for each skill for each language. Finally, in preparation for empirical piloting, ADT Team leaders were requested to select the best 300 items from their own item pool for the first round of piloting.

b) Piloting and calibration

In the second phase of the Project, arrangements were made for the computer-based piloting of items using the specially developed Pilot Tool software also developed by the Project (in the first phase of the Project, Finnish had been piloted in paper and pencil form, in order to try out both the overall pilot design and the calibration analyses). Initially a piloting design was implemented where 12 booklets would be piloted, containing 30 items from a skill (Reading, Writing, Listening) plus 20 items from a language aspect (Structures, Vocabulary), with each item appearing in two separate booklets, as well as 150 items from the Vocabulary Size Placement Test (VSPT) and between 30 and 40 self-assessment (SA) statements. Booklets were assigned randomly to pilot test-takers. This design would enable IRT-based item calibrations

to take place, using the computer program OPLM, once 600 learners had taken a pilot test (implying 100 responses per item) and a second calibration was to take place once 1200 users had responded (200 responses per item).

However, once sufficient numbers of responses to the VSPT and SAs had been gathered to enable calibrations, the pilot booklet design was modified to reduce the number of VSPT items being tested to 99 and the SA statements to 18 per skill. This allowed more test items to be included in each booklet, and only 450 test-takers would be required before an initial item calibration was possible. To date, initial calibrations have been possible for Finnish, English, French, German and Spanish, and Italian, Swedish and Dutch are getting close to the target number. English has undergone a second calibration, whose results were remarkably similar to the first round, suggesting a second calibration may not be necessary for other languages.

Standard Setting

Two different methods of standard setting were developed in DIALANG, and the Project developed guidelines which were subsequently published in the Manual and the Supplement. Details are available in Alderson (2005). In essence, the first method was a modified Angoff procedure, where probablity judgements by up to 12 experts were replaced by Yes-No decisions (numbers of judges varied by language and skill).

"The task for the judge is to state for each item in the test

At what CEF level can a test taker already answer the following item correctly?"
...
In practice the procedure follows a pattern also presented in the Manual:

- Familiarisation with the CEF
- Training (with feedback) on how to judge ... new items ... in terms of CEF levels;
- Judgement of items to be related to the CEF

(Manual, page 91)

This method was applied to Finnish, English and Spanish, and detailed results are reported in Alderson (2005).

Partly because of the lack of an adequate method of judging the reliability and consistency of judges, a second process was developed which involved experts in making two sets of judgements. The first round required them to place items into piles, representing the different levels of the CEFR, starting with A1, in answer to the question:

"Is it reasonable to require that a learner at level A1 gets this item right?"

If yes, then the item belongs to A1. If not, then it is a candidate for the next level up. Once piles have been created, they are checked for internal consistency and items may be moved if necessary.

After a period of time, minimally two to three hours, the process is repeated, thereby enabling intra-rater consistency to be calculated alongside the inter-rater agreement. Detailed results are available in Alderson (2005). Only reliable judges are retained in the subsequent analysis.

Finally, once item calibration details are available, the empirical results are combined with the judged CEFR level for each item, in a computer program specially developed by the Project.

References

Alderson, J. C. (2005) *Diagnosing foreign language proficiency: the interface between learning and assessment*. London: Continuum Books

Council of Europe (2003) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF). Manual: Preliminary Pilot Version*. Strasbourg: Language Policy Division, Council of Europe.

Information about the Dialang Items

The items presented in this section are taken from Version 1 of the freely available English test and represent a selection across the different CEFR levels. Each item is accompanied by a summary of its content taken from the Dutch CEF Grid (www.ling.lancs.ac.uk/cefgrid), itself closely based on the CEFR. In addition each item is accompanied by information on its statistical properties.

Statistical information

The calibration and standard setting processes described in the introduction to Dialang resulted in information about the properties of each item in terms of classical test theory - in particular, item facilities and discriminations. In addition, IRT analyses resulted in estimations of the ability of learners getting a given score, in terms of theta estimates of ability. Cut-offs on this theta scale are based on the standard-setting judgements, and an item's logit value can be related to the logit values for ability levels.

Guideline To Information On DIALANG items

Below follows an explanation of the statistical and other information accompanying the selected DIALANG items:

— the Framework level of the item is indicated immediately before the screenshot of the item
— the subskill of listening that the item was designed to measure is indicated immediately after the screenshot; DIALANG listening items aim at tapping one of these three skills:
  o Identifying the main idea(s) or information in the spoken discourse.
  o Listening intensively for specific detail.
  o Inferencing; i.e., the ability to infer meaning on the basis of spoken discourse.

The table under 'Information from piloting' describes the results for the booklets where the item was piloted.
— booklet number = the number of the pilot booklet in which the item was piloted (each DIALANG item appeared in two different booklets in the pilot design)
— n size = the number of pilot test takers for the booklet in question
— Facility = percent of pilot test takers who replied the item correctly
— Rpbis (unweighted) = point-biserial correlation between the item and the other listening items in the pilot booklet when all items have equal weight
— Rpbis (weighted) = point-biserial correlation between the item and the other listening items in the pilot booklet when the items are weighted by their discrimination index ('A' in the table under 'Results of calibration')

The table 'Results of calibration' display the key statistical information for the item from the OPLM analysis.

nr = number of the item in the OPLM analysis
label = label of the item (which in fact is also the item ID in the DIALANG system)
A = discrimination index
B = theta value of the item (which indicates its difficulty / position on the scale after the items have been scaled; the cut-offs for the Framework levels are expressed on this same scale)
SE(B) = standard error of the theta value
S = S test; in this the respondents are grouped into two or more (max 8) groups of equal size and the ability of the item to discriminate the groups is analysed
DF = degrees of freedom in the S analysis
P = probability of the S test (this in fact indicates whether the item studied fits the overall model)
M, M2, M3 = other indicators of the model fit for the item

— M = in this analysis, two groups are formed of the test takers and it is analysed whether the item can discriminate the two groups (the low group are those whose probability to get the item right is at most .4, whereas in the high group the probability is .6 or higher)
— M2 = in this analysis, respondents are split into two halves according to their overall score in the test / booklet
— M3 = in this analysis, the respondents are grouped into three groups of equal size, according to their test score

P or M-value of 99.999 indicates that a proper value cannot be computed because the respondents cannot be divided into equal groups; this occurs with very easy or very difficult items. For the same reason, the S test value can sometimes be .000.
An M value of the size of + or – 2 or above indicates that there may be a problem in the item (it cannot properly discriminate test takers at different ability levels)

Analysis of item content

The items on this CD have been analysed using the Dutch CEF Grid produced by the Dutch CEF Construct Project. This Project aimed to help test developers and other language educationists construct or relate test items to the CEFR. (The Final Report of the Project has been included on this CD for ease of reference.) One major outcome of the Project was an Internet-based Grid which can be used to help characterise reading and listening texts, items and tasks. Application of the Grid results in summary tables and the Grid has been used with the DIALANG items to produce the summary tables below. (Readers of this Users' Guide can access the Grid at:
www.ling.lancs.ac.uk/cefgrid
and can use it in the content analysis of their own texts, items and tasks.)

Dialang German Reading Comprehension Items

General information

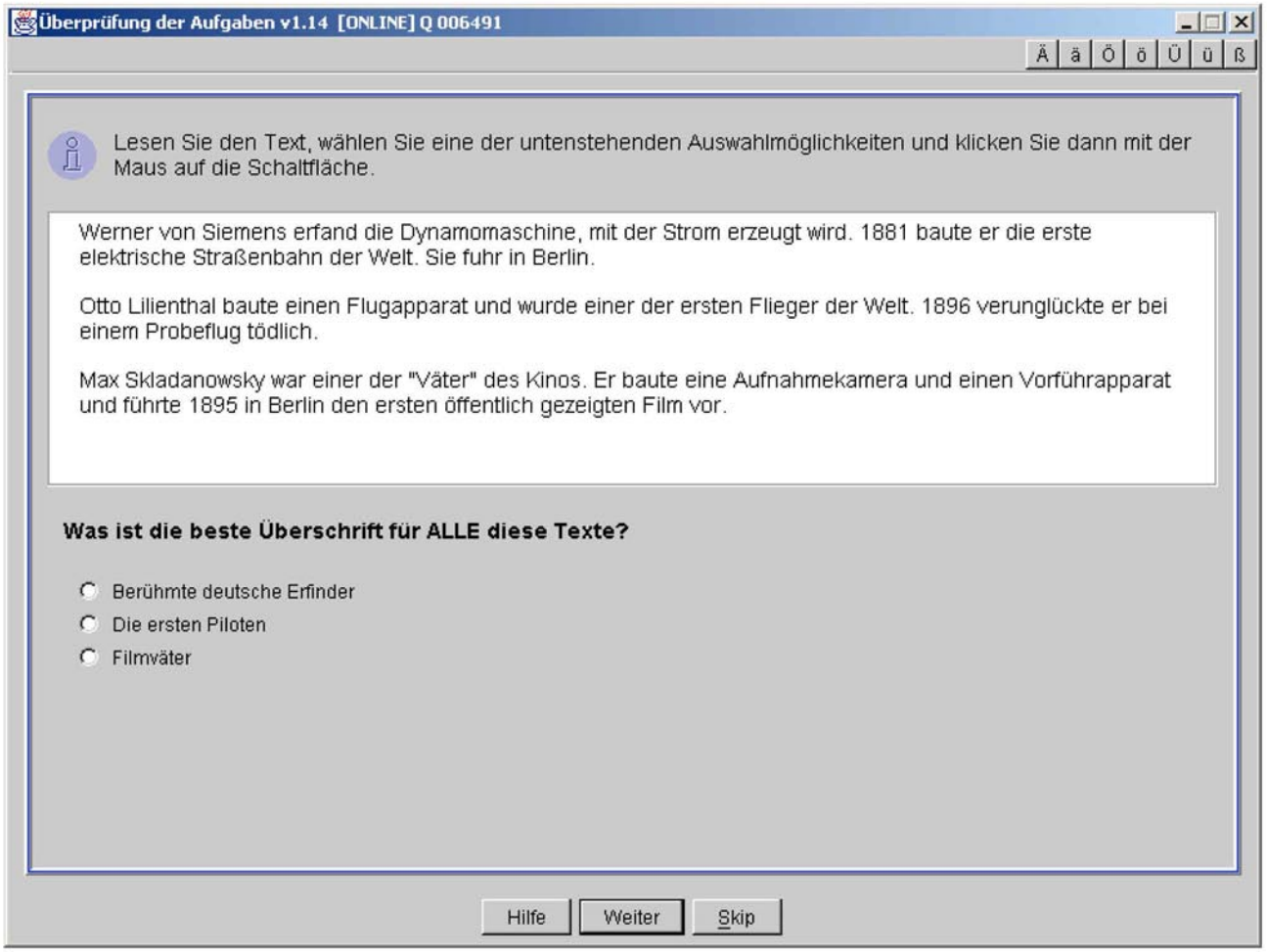
There were 137 – 143 pilot test takers per each reading item in the OPLM analyses. Overall data – model fit was satisfactory (p = .95).

The cut-offs on the theta scale for the CoE levels are based on standard setting procedures where 7 experts judged the items; the standard setting data were analysed with a special programme designed by Norman Verhelst at CITO.

Cut-off points for the Framework levels for the DIALANG German reading items:

under A1: theta under -.89
A1: theta between -.89 and -.43
A2: theta between -.43 and +.04
B1: theta between +.04 and +.54
B2: theta between +.54 and +1.19
C1: theta between +1.19 and +2.39
C2: theta above +2.39

A1


Überprüfung der Aufgaben v1.14 [ONLINE] Q 006491
Ä ä Ö ö Ü ü ß
Lesen Sie den Text, wählen Sie eine der untenstehenden Auswahlmöglichkeiten und klicken Sie dann mit der Maus auf die Schaltfläche.
Werner von Siemens erfand die Dynamomaschine, mit der Strom erzeugt wird. 1881 baute er die erste elektrische Straßenbahn der Welt. Sie fuhr in Berlin.
Otto Lilienthal baute einen Flugapparat und wurde einer der ersten Flieger der Welt. 1896 verunglückte er bei einem Probeflug tödlich.
Max Skladanowsky war einer der "Väter" des Kinos. Er baute eine Aufnahmekamera und einen Vorführapparat und führte 1895 in Berlin den ersten öffentlich gezeigten Film vor.
Was ist die beste Überschrift für ALLE diese Texte?
Berühmte deutsche Erfinder
Die ersten Piloten
Filmväter
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	6491
1. Text Source:	Reference books
2. Authenticity:	Abridged/adapted/simplified
3. Discourse type:	Mainly Expository
4. Domain:	Public
5. Topic:	15. Other
Specify other topic:	Invention
6. Nature of Content:	mostly concrete content
7. Text Length:	70
8. Vocabulary:	rather extended
9. Grammar:	mainly simple structures
Comprehensible by learner at CEF level:	A2/B1

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Evaluate
15. Operations:	Explicit
15. Operations:	Main idea/gist
Item Level Estimated:	A2/B1

Comment: The expert judges doing the content analysis overestimated the difficulty level of the item.

Statistical analysis

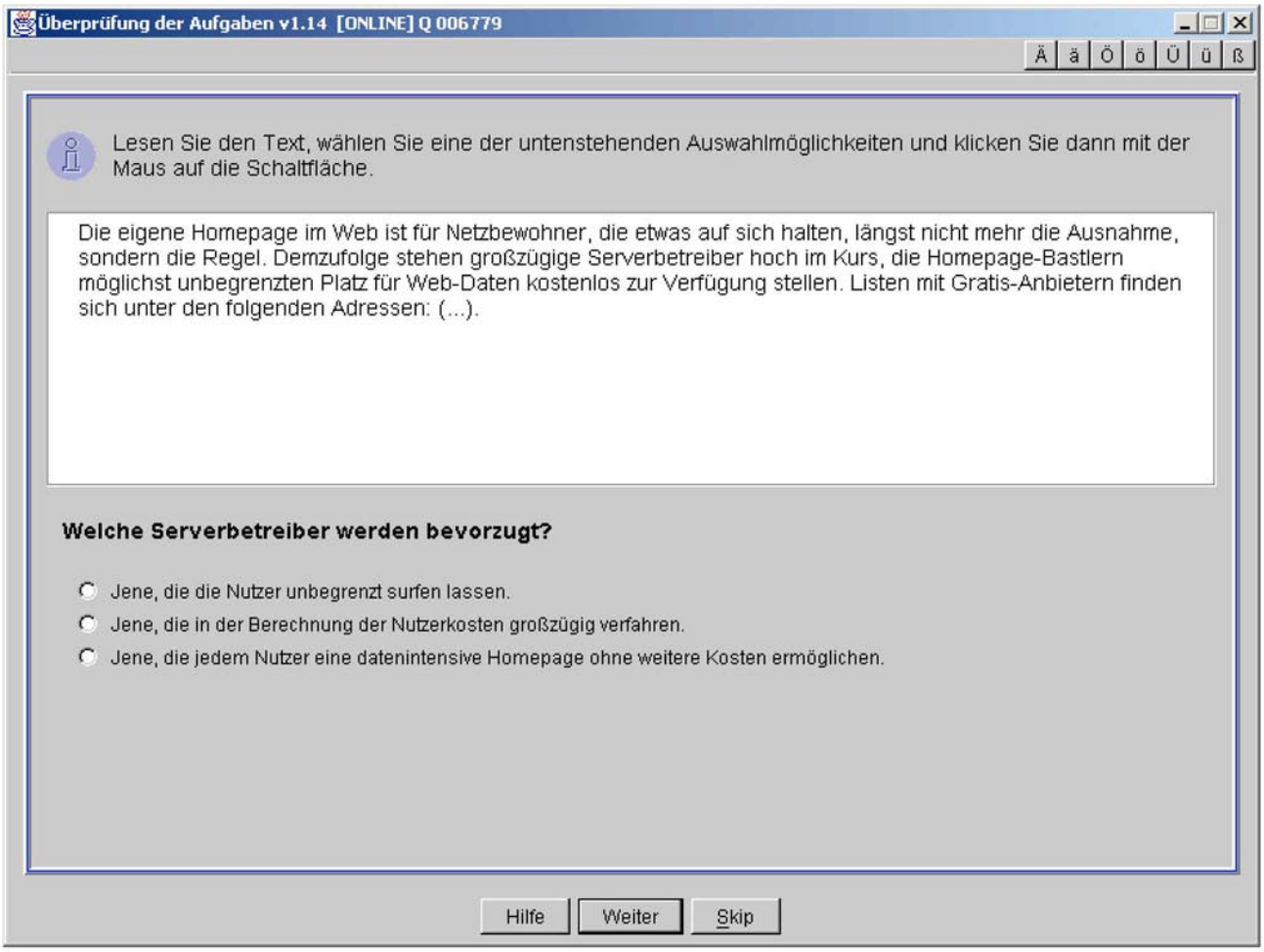
Information from piloting:

Booklet number	70n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
6	70	.886	.519	.549
7	70	.943	.364	.395

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
79 R006491	4	-.740	.091	.000	0	99.999	.521	-.821	-1.455

A2


Überprüfung der Aufgaben v1.14 [ONLINE] Q 006779
Ä ä Ö ö Ü ü ß
Lesen Sie den Text, wählen Sie eine der untenstehenden Auswahlmöglichkeiten und klicken Sie dann mit der Maus auf die Schaltfläche.
Die eigene Homepage im Web ist für Netzbewohner, die etwas auf sich halten, längst nicht mehr die Ausnahme, sondern die Regel. Demzufolge stehen großzügige Serverbetreiber hoch im Kurs, die Homepage-Bastlern möglichst unbegrenzten Platz für Web-Daten kostenlos zur Verfügung stellen. Listen mit Gratis-Anbietern finden sich unter den folgenden Adressen: (...).
Welche Serverbetreiber werden bevorzugt?
Jene, die die Nutzer unbegrenzt surfen lassen.
Jene, die in der Berechnung der Nutzerkosten großzügig verfahren.
Jene, die jedem Nutzer eine datenintensive Homepage ohne weitere Kosten ermöglichen.
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	6779
1. Text Source:	Newspapers
2. Authenticity:	Authentic
3. Discourse type:	Mainly Descriptive
4. Domain:	Public
5. Topic:	15. Other
Specify other topic:	Internet
6. Nature of Content:	mostly concrete content
7. Text Length:	70
8. Vocabulary:	rather extended
9. Grammar:	Limited range of complex structures
Comprehensible by learner at CEF level:	B1

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Recognise and Retrieve
15. Operations:	Explicit
15. Operations:	Detail
Item Level Estimated:	B1

Comment: The expert judges doing the content analysis overestimated the difficulty level of the item.

Statistical analysis

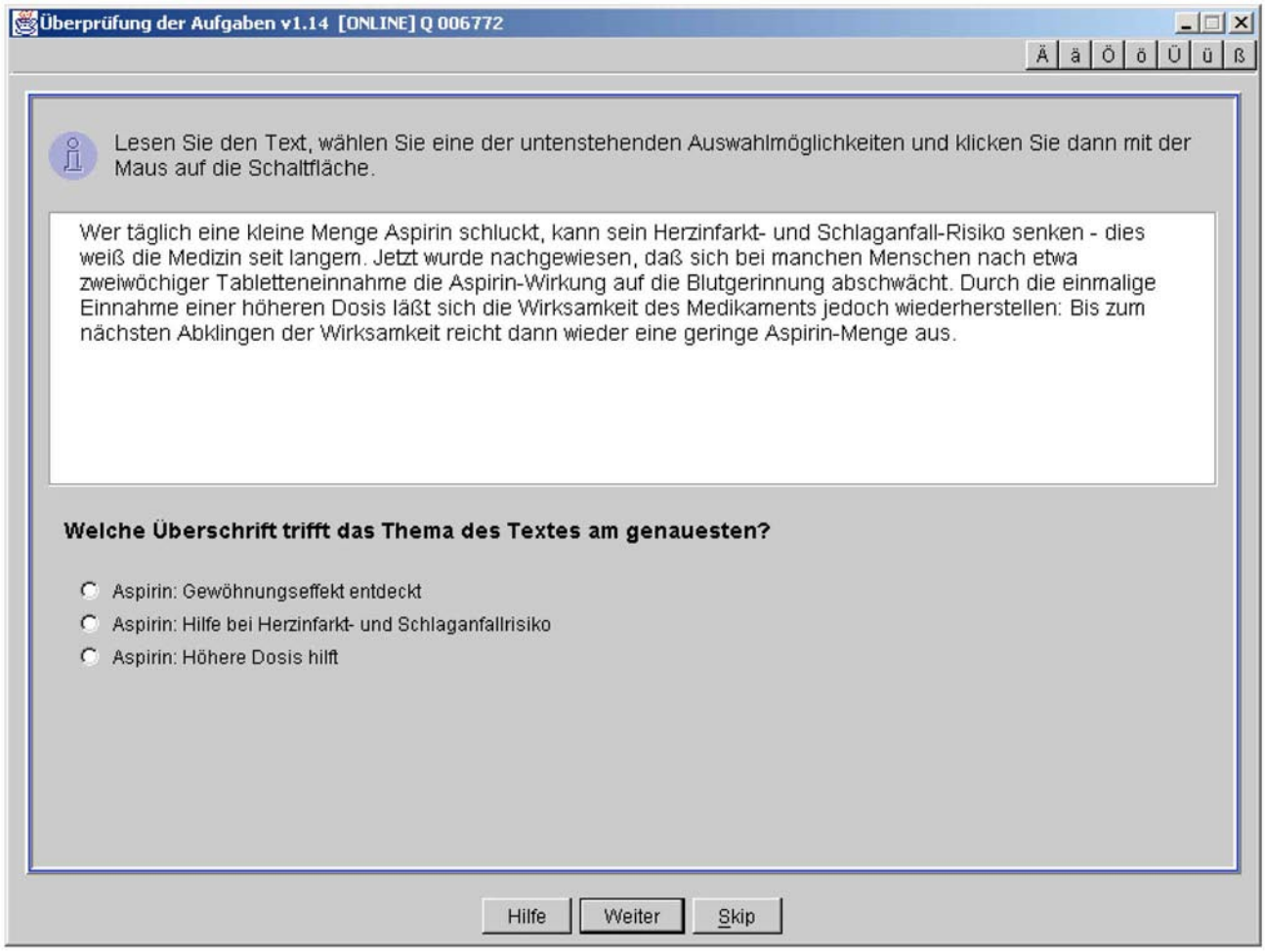
Information from piloting:

Booklet number	n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
6	70	.714	.493	.480
7	70	.729	.580	.582

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
85 R006779	3	-.296	.052	6.927	5	.226	-.544	-1.984	-.827

B1


Überprüfung der Aufgaben v1.14 [ONLINE] Q 006772
Ä ä Ö ö Ü ü ß
Lesen Sie den Text, wählen Sie eine der untenstehenden Auswahlmöglichkeiten und klicken Sie dann mit der Maus auf die Schaltfläche.
Wer täglich eine kleine Menge Aspirin schluckt, kann sein Herzinfarkt- und Schlaganfall-Risiko senken - dies weiß die Medizin seit langem. Jetzt wurde nachgewiesen, daß sich bei manchen Menschen nach etwa zweiwöchiger Tabletteneinnahme die Aspirin-Wirkung auf die Blutgerinnung abschwächt. Durch die einmalige Einnahme einer höheren Dosis läßt sich die Wirksamkeit des Medikaments jedoch wiederherstellen: Bis zum nächsten Abklingen der Wirksamkeit reicht dann wieder eine geringe Aspirin-Menge aus.
Welche Überschrift trifft das Thema des Textes am genauesten?
Aspirin: Gewöhnungseffekt entdeckt
Aspirin: Hilfe bei Herzinfarkt- und Schlaganfallrisiko
Aspirin: Höhere Dosis hilft
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	6772
1. Text Source:	Newspapers
2. Authenticity:	Authentic
3. Discourse type:	Mainly Expository
4. Domain:	Public
5. Topic:	7. Health and bodycare
6. Nature of Content:	mostly concrete content
7. Text Length:	68
8. Vocabulary:	rather extended
9. Grammar:	Limited range of complex structures
Comprehensible by learner at CEF level:	B1/B2

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Evaluate
15. Operations:	Implicit
15. Operations:	Main idea/gist
Item Level Estimated:	B1/B2

Statistical analysis

Information from piloting:

Booklet number	n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
8	73	.370	.430	.430
9	67	.507	.539	.554

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
110 R006772	3	.287	.067	2.495	3	.476	-.696	-.486	-.286

Dialang German Listening Comprehension Items

General information

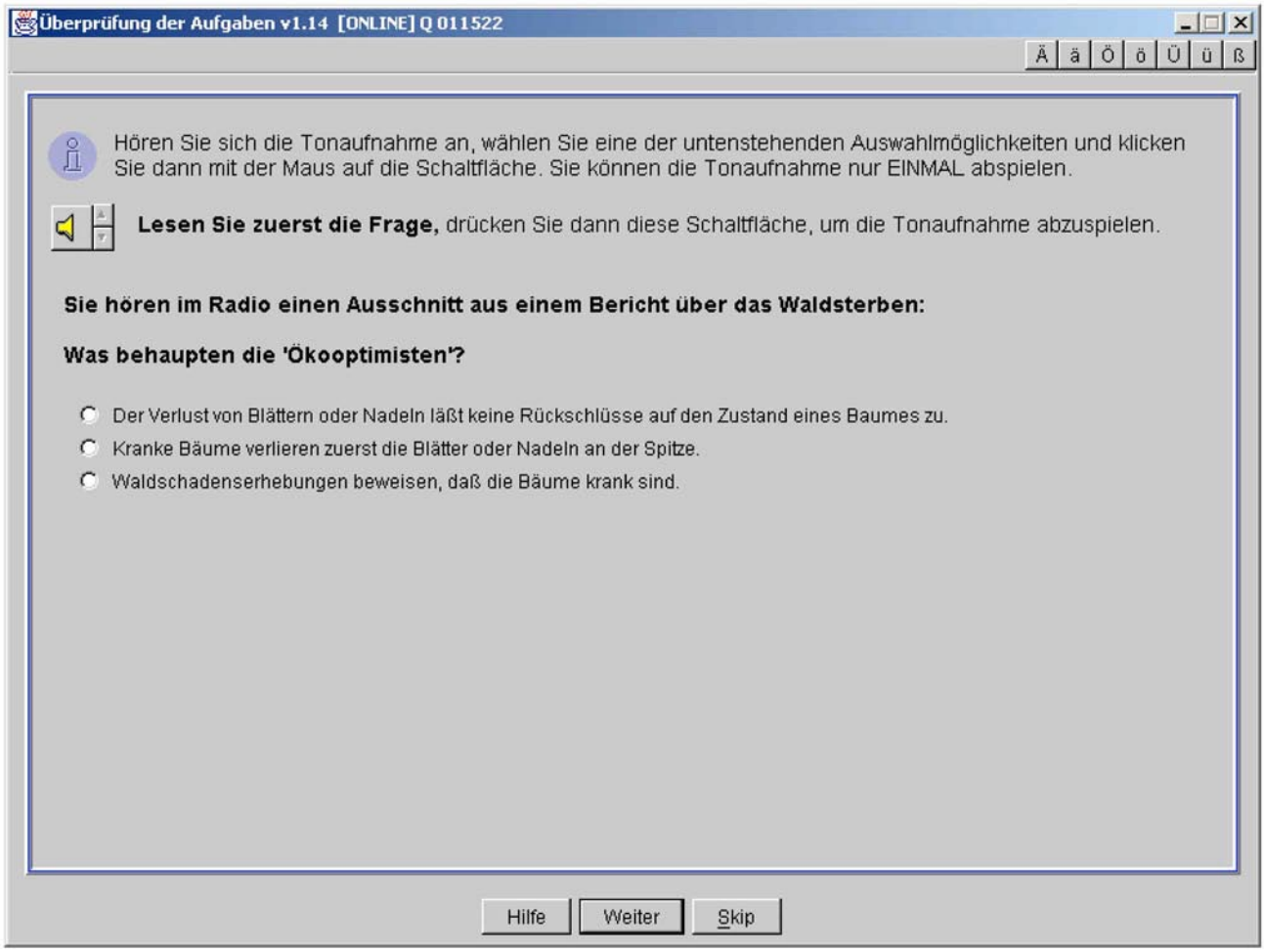
There were 140 – 183 pilot test takers per each listening item in the OPLM analyses. Overall data – model fit was satisfactory (p = .20).

The cut-offs on the theta scale for the CoE levels are based on standard setting procedures where 7 experts judged the items; the standard setting data were analysed with a special programme designed by Norman Verhelst at CITO.

Cut-off points for the Framework levels for the DIALANG German LISTENING items:

under A1: theta under -.41
A1: theta between -.41 and -.08
A2: theta between -.08 and +.16
B1: theta between +.16 and +.38
B2: theta between +.38 and +.68
C1: theta between +.68 and +1.56
C2: theta above +1.56

A2


Überprüfung der Aufgaben v1.14 [ONLINE] Q 011522
Ä ä Ö ö Ü ü ß
Hören Sie sich die Tonaufnahme an, wählen Sie eine der untenstehenden Auswahlmöglichkeiten und klicken Sie dann mit der Maus auf die Schaltfläche. Sie können die Tonaufnahme nur EINMAL abspielen.
Lesen Sie zuerst die Frage, drücken Sie dann diese Schaltfläche, um die Tonaufnahme abzuspielen.
Sie hören im Radio einen Ausschnitt aus einem Bericht über das Waldsterben:
Was behaupten die 'Ökooptimisten'?
Der Verlust von Blättern oder Nadeln läßt keine Rückschlüsse auf den Zustand eines Baumes zu.
Kranke Bäume verlieren zuerst die Blätter oder Nadeln an der Spitze.
Waldschadenserhebungen beweisen, daß die Bäume krank sind.
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	11522
1. Text Source:	TV/ radio documentaries
2. Authenticity:	Authentic
3. Discourse type:	Mainly Argumentative
4. Domain:	Public
5. Topic:	15. Other
Specify other topic:	Ecology
6. Nature of Content:	fairly extensive abstract content
7. Text Length:	18secs
8. Vocabulary:	rather extended
9. Grammar:	Limited range of complex structures
10. Text Speed:	normal
11. No of participants:	one
12. Accent Standard:	Standard pronunciation
13. Clarity of Articulation:	normally articulated
14. How often played:	played once
Comprehensible by learner at CEF level:	B1

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Make inferences
15. Operations:	Explicit
15. Operations:	Main idea/gist
15. Operations:	Conclusion
Item Level Estimated:	B1

Comment: The expert judges doing the content analysis overestimated the difficulty level of the item.

Statistical analysis

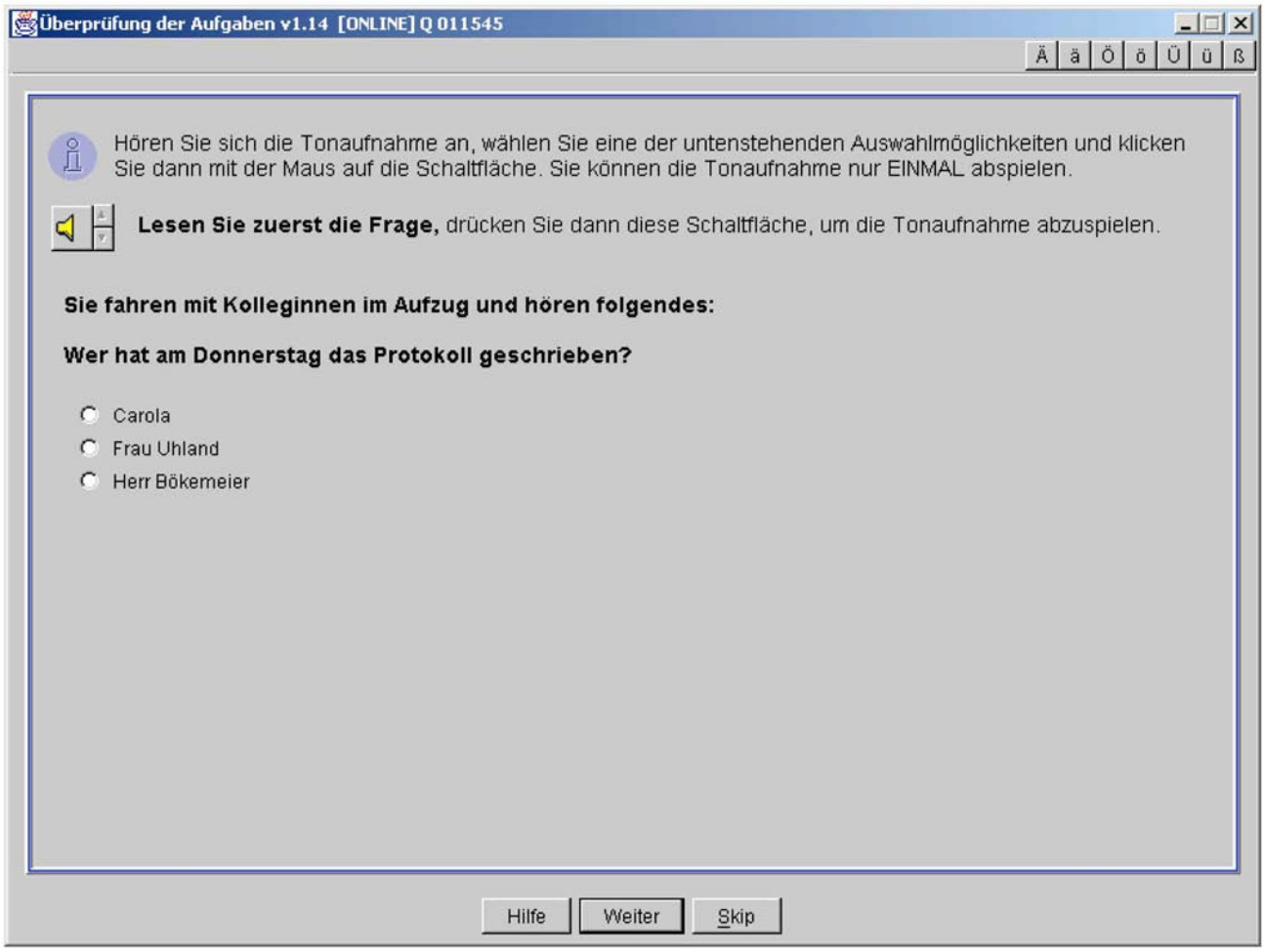
Information from piloting:

Booklet number	n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
1	82	.585	.635	.664
2	101	.693	.426	.427

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
24 L011522	3	.145	.060	1.893	3	.595	-1.065	-1.099	-.496

B1


Überprüfung der Aufgaben v1.14 [ONLINE] Q 011545
Ä ä Ö ö Ü ü ß
Hören Sie sich die Tonaufnahme an, wählen Sie eine der untenstehenden Auswahlmöglichkeiten und klicken Sie dann mit der Maus auf die Schaltfläche. Sie können die Tonaufnahme nur EINMAL abspielen.
Lesen Sie zuerst die Frage, drücken Sie dann diese Schaltfläche, um die Tonaufnahme abzuspielen.
Sie fahren mit Kolleginnen im Aufzug und hören folgendes:
Wer hat am Donnerstag das Protokoll geschrieben?
Carola
Frau Uhland
Herr Bökemeier
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	11545
1. Text Source:	Interpersonal dialogues and conversations
2. Authenticity:	Authentic
3. Discourse type:	Mainly Narrative
4. Domain:	Occupational
5. Topic:	15. Other
Specify other topic:	Work
6. Nature of Content:	fairly extensive abstract content
7. Text Length:	31secs
8. Vocabulary:	rather extended
9. Grammar:	mainly simple structures
10. Text Speed:	fast
11. No of participants:	one
12. Accent Standard:	Standard pronunciation
13. Clarity of Articulation:	normally articulated
14. How often played:	played once
Comprehensible by learner at CEF level:	B1/B2

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Recognise and Retrieve
15. Operations:	Explicit
15. Operations:	Detail
Item Level Estimated:	B1/B2

Statistical analysis

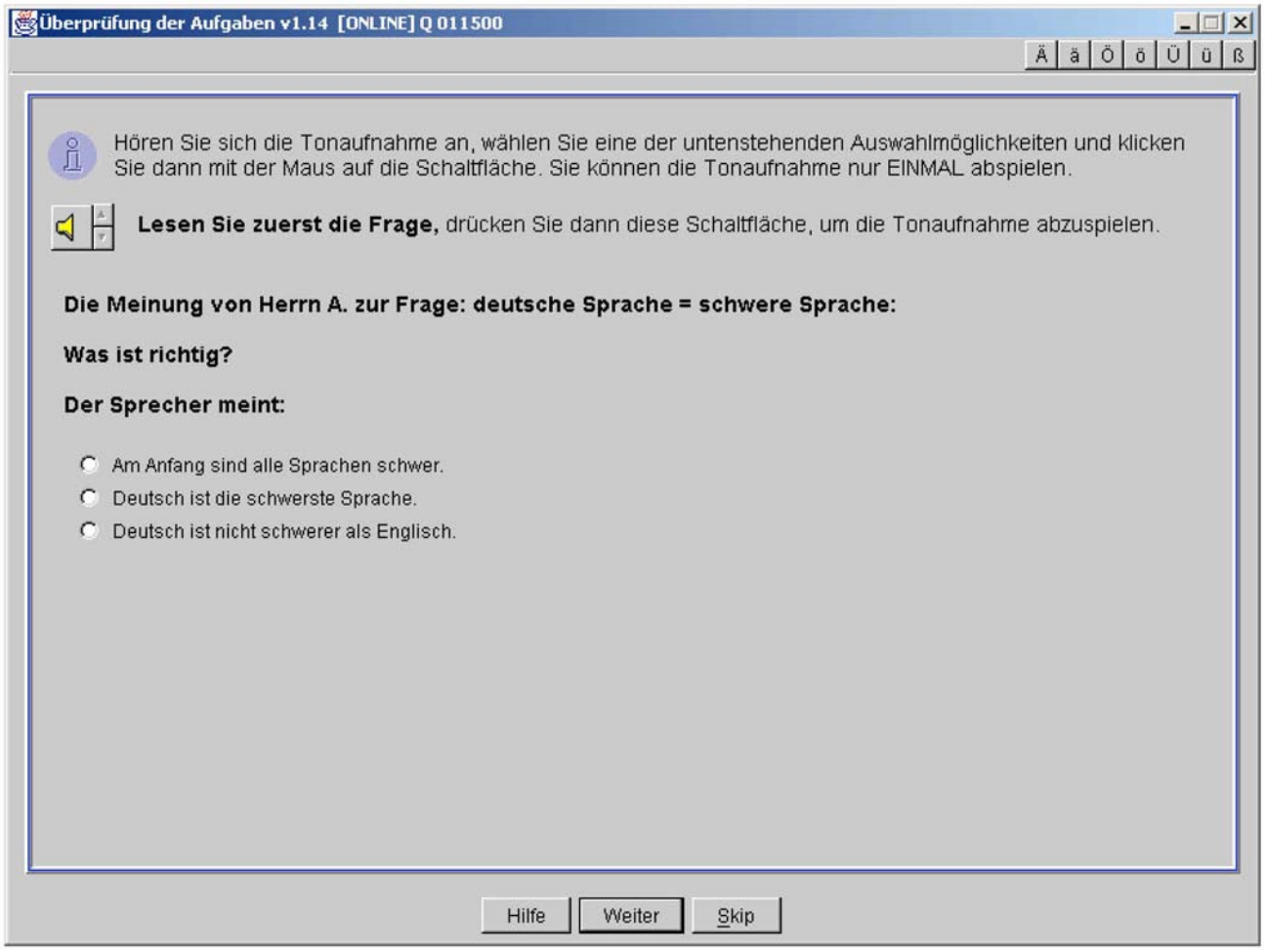
Information from piloting:

Booklet number	n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
1	82	.585	.601	.645
4	64	.641	.485	.489

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
11 L011545	3	.184	.068	4.705	3	.195	-.602	-1.255	-.816

B2


Überprüfung der Aufgaben v1.14 [ONLINE] Q 011500
Ä ä Ö ö Ü ü ß
Hören Sie sich die Tonaufnahme an, wählen Sie eine der untenstehenden Auswahlmöglichkeiten und klicken Sie dann mit der Maus auf die Schaltfläche. Sie können die Tonaufnahme nur EINMAL abspielen.
Lesen Sie zuerst die Frage, drücken Sie dann diese Schaltfläche, um die Tonaufnahme abzuspielen.
Die Meinung von Herrn A. zur Frage: deutsche Sprache = schwere Sprache:
Was ist richtig?
Der Sprecher meint:
Am Anfang sind alle Sprachen schwer.
Deutsch ist die schwerste Sprache.
Deutsch ist nicht schwerer als Englisch.
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	11500
1. Text Source:	Public speeches, lectures, presentations, sermons
2. Authenticity:	Authentic
3. Discourse type:	Mainly Argumentative
4. Domain:	Public
4. Domain:	Educational
5. Topic:	13. Language
6. Nature of Content:	fairly extensive abstract content
7. Text Length:	23secs
8. Vocabulary:	mostly frequent vocabulary
9. Grammar:	mainly simple structures
10. Text Speed:	normal
11. No of participants:	one
12. Accent Standard:	Standard pronunciation
13. Clarity of Articulation:	normally articulated
14. How often played:	played once
Comprehensible by learner at CEF level:	B1

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Recognise and Retrieve
15. Operations:	Explicit
15. Operations:	Main idea/gist
15. Operations:	Opinion
Item Level Estimated:	B1

Comment: The expert judges doing the content analysis underestimated the difficulty level of the item.

Statistical analysis

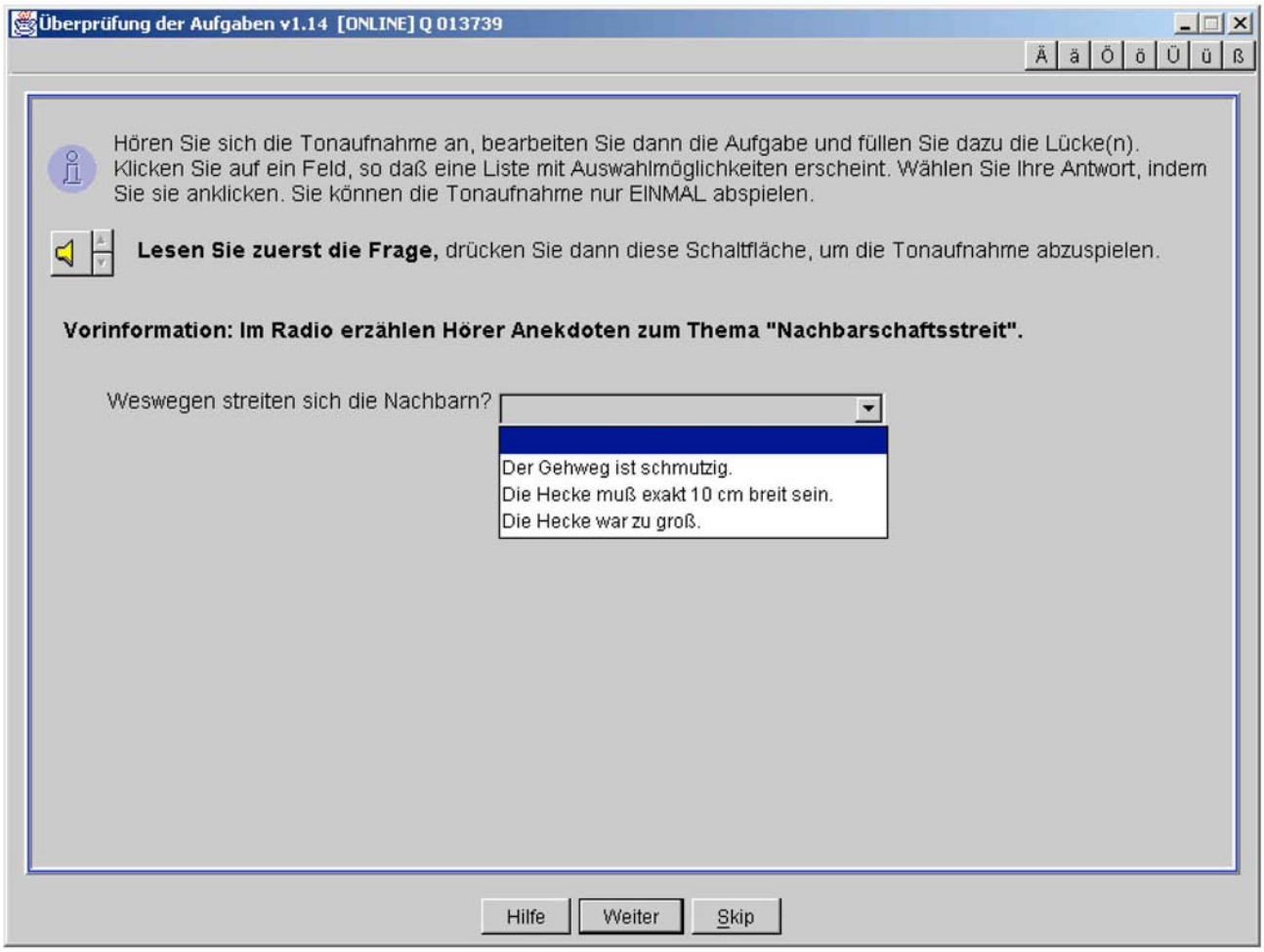
Information from piloting:

Booklet number	n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
1	82	.402	.652	.655
4	64	.422	.432	.461

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
8 L011500	4	.526	.057	.289	2	.865	.816	-.013	1.006

B2


Überprüfung der Aufgaben v1.14 [ONLINE] Q 013739
Ä ä Ö ö Ü ü ß
Hören Sie sich die Tonaufnahme an, bearbeiten Sie dann die Aufgabe und füllen Sie dazu die Lücke(n). Klicken Sie auf ein Feld, so daß eine Liste mit Auswahlmöglichkeiten erscheint. Wählen Sie Ihre Antwort, indem Sie sie anklicken. Sie können die Tonaufnahme nur EINMAL abspielen.
Lesen Sie zuerst die Frage, drücken Sie dann diese Schaltfläche, um die Tonaufnahme abzuspielen.
Vorinformation: Im Radio erzählen Hörer Anekdoten zum Thema "Nachbarschaftsstreit".
Weswegen streiten sich die Nachbarn?
Der Gehweg ist schmutzig.
Die Hecke muß exakt 10 cm breit sein.
Die Hecke war zu groß.
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	13739
1. Text Source:	Radio phone-in
2. Authenticity:	Authentic
3. Discourse type:	Mainly Narrative
4. Domain:	Public
5. Topic:	6. Relations with other people
6. Nature of Content:	mostly concrete content
7. Text Length:	21secs
8. Vocabulary:	mostly frequent vocabulary
9. Grammar:	mainly simple structures
10. Text Speed:	normal
11. No of participants:	one
12. Accent Standard:	Standard pronunciation
13. Clarity of Articulation:	clearly articulated
14. How often played:	played once
Comprehensible by learner at CEF level:	B2

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Make inferences
15. Operations:	Explicit
15. Operations:	Main idea/gist
Item Level Estimated:	B2

Statistical analysis

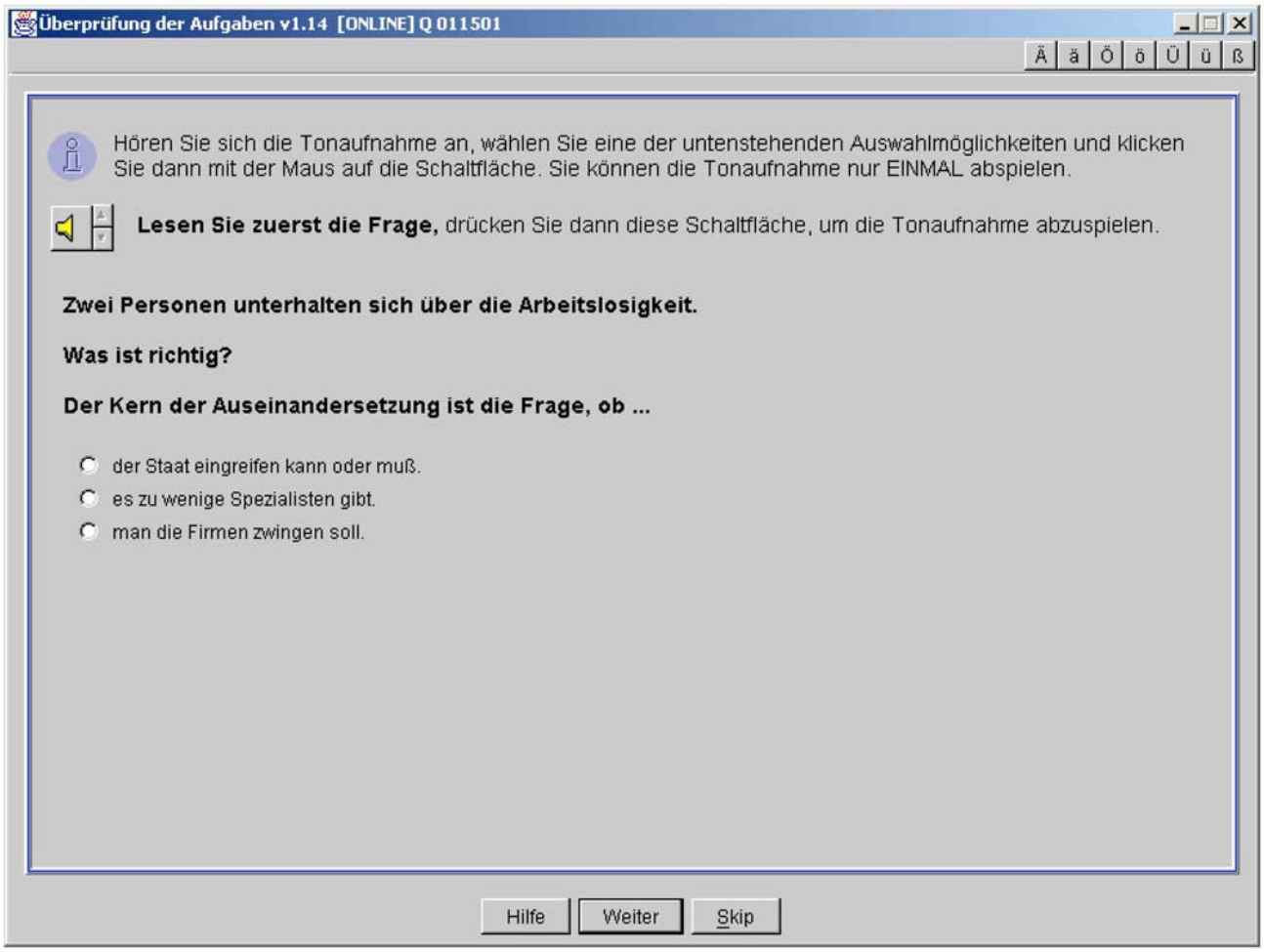
Information from piloting:

Booklet number	n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
1	82	.439	.653	.677
4	64	.484	.474	.525

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
12 L013739	4	.451	.056	1.146	2	.564	.166	.612	.708

C1


Überprüfung der Aufgaben v1.14 [ONLINE] Q 011501
Ä ä Ö ö Ü ü ß
Hören Sie sich die Tonaufnahme an, wählen Sie eine der untenstehenden Auswahlmöglichkeiten und klicken Sie dann mit der Maus auf die Schaltfläche. Sie können die Tonaufnahme nur EINMAL abspielen.
Lesen Sie zuerst die Frage, drücken Sie dann diese Schaltfläche, um die Tonaufnahme abzuspielen.
Zwei Personen unterhalten sich über die Arbeitslosigkeit.
Was ist richtig?
Der Kern der Auseinandersetzung ist die Frage, ob ...
der Staat eingreifen kann oder muß.
es zu wenige Spezialisten gibt.
man die Firmen zwingen soll.
Hilfe
Weiter
Skip

Content analysis

Text Characteristics:

Task:	11501
1. Text Source:	Interpersonal dialogues and conversations
2. Authenticity:	Authentic
3. Discourse type:	Mainly Argumentative
4. Domain:	Personal
4. Domain:	Occupational
5. Topic:	15. Other
Specify other topic:	Employment
6. Nature of Content:	fairly extensive abstract content
7. Text Length:	36secs
8. Vocabulary:	rather extended
9. Grammar:	Limited range of complex structures
10. Text Speed:	normal
11. No of participants:	two
12. Accent Standard:	Standard pronunciation
13. Clarity of Articulation:	normally articulated
14. How often played:	played once
Comprehensible by learner at CEF level:	B2

Item Characteristics:

Item:	1
14. Item Type:	1. Multiple choice
15. Operations:	Make inferences
15. Operations:	Explicit
15. Operations:	Main idea/gist
Item Level Estimated:	B2

Comment: The expert judges doing the content analysis underestimated the difficulty level of the item.

Statistical analysis

Information from piloting:

Booklet number	n size	Facility	Rpbis (unweighted)	Rpbis (weighted)
1	82	.354	.237	.164
2	64	.453	.159	.097

Results of calibration:

nr label	A	B	SE(B)	S	DF	P	M	M2	M3
6 L011501	1 .840	.173	.783	3	.854	.816	.788	.811	