TNO Quality of Life

Wassenaarseweg 56
P.O. Box 2215
2301 CE Leiden
The Netherlands

www.tno.nl

T +31 71 518 18 18
F +31 71 518 19 20

TNO report

KvL/P&Z 2006.048

An Item Response Theory driven Computerized Adaptive Test as an effective and efficient tool to improve the identification by Preventive Child Health Care of children with psychosocial problems

Date June 2006

Author(s) A.G.C. Vogels
G.W. Jacobusse
R.K.R.R. Chablani
A.J.M. Rovekamp
M.J.C. Hilgersom
S.A. Reijneveld

Number of pages 65
Number of appendices 1

All rights reserved.
No part of this publication may be reproduced and/or published by print, photoprint, microfilm or any other means without the previous written consent of TNO.

In case this report was drafted on instructions, the rights and obligations of contracting parties are subject to either the Standard Conditions for Research Instructions given to TNO, or the relevant agreement concluded between the contracting parties. Submitting the report for inspection to parties who have a direct interest is permitted.

© 2006 TNO

Preface

This report describes the results of the project 'A better identification of 7-12 years old children with psychosocial problems by means of an intelligent computer based questionnaire, based on Item Response Theory'. The reasons why this study was started was, shortly, the inherent psychometric weaknesses of short paper and pencil questionnaires as they are currently being used in Preventive Child Health Care, in combination with the fact that in other fields Item Response Theory (IRT) based Computerized adaptive Tests (CAT) allow for very short, yet high quality measurement procedures.

The project took place in two stages. In the first stage a secondary analysis was done on data collected in an earlier study (Vogels et al, 2005). Aim of the analyses was to determine if items of four questionnaires on psychosocial problems among children aged seven till twelve were suited for a Computerized adaptive Test (CAT) based on Item Response Theory (IRT). This stage of the study was approved and supported by ZonMw.

As the results of the first part of the study were positive, TNO Quality of Life decided to start the second stage of this study, in which a working prototype for a CAT on psychosocial problems was developed.

The first part of this report describes the activities and the results of the first stage of the study. The second part describes the development of the working prototype.

Based on the findings of this study TNO Quality of Life is convinced that IRT based CAT is a very promising tool, which allows for an efficient and precise identification of children who need further attention because of likely psychosocial problems. Of course, a secondary analysis can not assess the usability in real life, nor can it assess the validity of the results when used in daily practice. Definite conclusions on usability and validity can only be based on an evaluation and validity study in real life.

This project focused on children aged seven till twelve years. The methods and tools we developed are applicable, however, on a much wider scale. They can be used for children in other age groups and for other measurements, if and when suitable items for such measurements are available. They could also be used to improve monitoring of individual children by increasing the frequency of the measurements without much extra costs and thereby collecting information on mental health as the child develops.

The very promising results of this project and the wider applicability of the tools we developed, call for a pilot implementation in which both the real life validity and usability can be tested.

Contents

1 Item Response Theory based Computerized adaptive Testing: could it improve the identification of children with psychosocial problems?

1.1 Introduction

Many children suffer from behavioural and emotional problems ({Brugman, 2001 419 /id;Kelleher, 2000 418 /id;Horowitz, 2001 153 /id}) and as such problems may seriously interfere with daily functioning, now and later in life{Moffitt T.E., 2001 558 /id;Verhulst, 1997 557 /id}. Yet many of these children remain untreated {Verhulst, 1997 541 /id}. Durlak and Wells showed that early identification does improve the prognosis of the children involved considerably {Durlak, 1997 466 /id}.

Community based preventive child health services (PCH), especially those working outreachingly, towards the population as a whole, are in a unique position to identify such problems as early as possible, and many of them try to do so. Research has shown, however, that early identification in PCH is often far from perfect. For example, Brugman showed that in Dutch PCH, about half of the children with a clinical CBCL Total Problem Score remain unnoticed when they are examined by a physician or nurse {Brugman, 2001 419 /id}. Other studies came to similar conclusions {Glascoe, 1999 249 /id;Lavigne, 1993 454 /id;Costello, 1990 412 /id;Reijneveld, 2004 404 /id}.

There are several possibilities to improve the identification of children with emotional and behavioural problems. Wiefferink et. al. showed that using clear protocols and extensive staff training can lead to a significant increase in the number of children with problems identified increased significantly and a decrease in the number of children incorrectly identified as having problems decreased {Wiefferink, 2006 566 /id}. Several studies also showed that using high quality questionnaires, to be filled in by parents, teachers or the children themselves can also help to improve the quality of early identification {Durlak, 1997 466 /id;Reijneveld, 2003 19 /id;Vogels, 2003 537 /id}. However, in PCH which try to target, outreachingly, the population as a whole, there is only limited time available for each individual child. This means that questionnaires that are practicable in such settings, have to be short, easy to administer and easy to score. And in traditional psychometrics, short questionnaires, unless they have a very narrow scope, tend to be less reliable and less valid than desirable. Identification of problems based on such questionnaires must then be error prone, resulting in too many false classifications.

Since the 1950's a series of new statistical models called RASCH or IRT (Item Response Theory) models have been developed which allow for short and efficient test procedures without compromising the test's results' accuracy. Originally limited to items with only two categories, the application of such models were limited mainly to the field of intelligence testing and the assessment of school achievements ({Wright, 1982 606 /id}). In the last decades more widely applicable models became available. This lead to IRT-based test procedures in the field of quality of life measurements ({Revicki, 1997 607 /id}). Very recently some publications are beginning to appear which try to apply such models on the assessment of mental health problems{Gardner, 2004 548 /id;Gardner, 2002 550 /id;Fliege, 2005 608 /id}.

In this study we want to assess whether these IRT models can also be used for a fast, short, yet high quality identification of children with emotional and behavioural problems in community based PCH.

1.2 Methods

1.2.1 General outline

We wanted to assess whether a test procedure based on IRT models allows for an accurate distinction between children with and without problems, using only a small number of questions. IRT based testing does so by adjusting questions to individual respondents. Each respondent has a certain amount of psychosocial problems. This amount can be thought of as a position (person's location) on a latent unidimensional scale. Not only respondents, but also items have a position on the latent scale (item location) {Wright, 1982 606 /id}. A good question for a person, that is a question giving distinctive information about this specific person, is a question with an item location that is close to the person's location. Each time a question is answered, the estimated person's location is updated and the best next item is selected, until a previously defined accuracy has been reached. In practice this process is only possible with the aid of computers: Computerized Adaptive Testing (CAT) {Wainer, 1990 604 /id}. For a CAT to be possible, the item locations need to be known in advance, before actual testing starts.

For the analyses we used an available data set of parents reporting on their children's problems (see below) on paper and pencil questionnaires. This sample of 2041 respondents was divided in two random sub samples: a calibration sample of 1650 and a validation sample of 391 respondents. In the first stage of the analysis, we used the calibration sample in order to assess the fit to the IRT model for items on questionnaires used to identify children with emotional and behavioural problems and in order to estimate the item locations, needed for an IRT based CAT. We also used this sample to determine a suitable cut-off point, comparable to a CBCL clinical cut-off point.

In the second stage, we used the validation sample in a simulation study in which CAT was used to select a subset of questions and determined whether this CAT succeeded in selecting the right questions to measure persons' location accurately. We used the answers on the selected items, as given on the paper and pencil questionnaires, as input in the simulated CAT assessment. The resulting estimated person's location was then compared to two criteria, having a clinical score on the Total problems Scale (TPS) of the Child Behavior Checklist (CBCL) and currently being under treatment for psychosocial problems.

The random division of the original sample into the calibration and validation sample ensured that we didn't overestimate accuracy by applying a model to the same data that were used to construct it.

1.2.2 Data collection, population and measures

We used a data set collected in an earlier study {Vogels, 2005 561 /id} containing information of parent reported problems of children aged seven till twelve. Data were collected in a two step procedure. In the first step nine regional PCH were found willing to participate in our study. Second, parents who were invited for a routine ehalth

examination of their child were asked to participate in the study to fill in some questionnaires about emotional and behavioural disorders of their child. Total response was 84%.

Data from 2041 parents were available. 51% of the children involved were girls. Mean age was 10.1 (sd=1.4). In 83% of all cases both parents were born in the Netherlands and in 6% at least one parent came from a non-OESO country or Turkey. Fourteen percent did not live in a two-parent family and only three percent lived in a family were none of the parents were unemployed. Due to the way data collection took place the sample may be considered as being largely representative for the population of this age group under care in Dutch PCH. Children from ethnic minorities and children from unemployed parents were underrepresented, however.

Each parent filled in the CBCL. The CBCL was developed originally by Achenbach {Achenbach, 2001 575 /id}. It is a 121 item questionnaire. In this study the Dutch version of the 1991 CBCL {Verhulst, 1996 536 /id;Verhulst, 1997 556 /id;Verhulst, 1992 572 /id} was used. Each items has three answering categories (not at all, a bit/sometimes, clearly/often).

In addition each parent answered one out of three questionnaires: the Pediatric Symptom Checklist ({Jellinek, 1979 1 /id;Jellinek, 1995 323 /id;Jellinek, 1988 545 /id}), the Strengths and Difficulties Questionnaire (SDQ){Bourdon, 2005 589 /id;Goodman, 2003 78 /id;Goodman, 1999 253 /id} or a newly developed Dutch questionnaire on psychosocial problems for children in primary education, the PSYBOBA{Eden van, 2003 533 /id}. This procedure lead to an incomplete data matrix: the data for the PSC, the SDQ and the PSYBOBA are each available in about one third of the sample.

The PSC consists of 35 items on problems. The PSC is used extensively in community based PCH in the USA. As no Dutch version was available, we translated the PSC using three independent translators and back translators and feedback from the original authors {Guillemin, 1995 542 /id;Guillemin, 1993 543 /id}. Several studies describe the validity and usability of the PSC in several settings {Anderson, 1999 6 /id;Jellinek, 1995 323 /id;Jellinek, 1988 545 /id;Jellinek, 1988 5 /id;Murphy, 1992 4 /id;Jutte, 2003 8 /id}. Each item has 3 answering categories (never, sometimes, always).

The SDQ was developed by Goodman et al. It contains 25 items relating to the child's problems or strong points in its functioning. Its validity and usability in several settings and several countries and cultures is widely documented {Bourdon, 2005 589 /id;Goodman, 2000 181 /id;Goodman, 1999 253 /id;Goodman, 2000 602 /id}. Each item has 3 answering categories (not true, more or less true, certainly true).

The PSYBOBA is a new Dutch instrument, designed specifically to be used in Dutch PCH's. It consists of 26 items relating to the child's problems. Some evidence for its validity and usability is available (van Eden, Vogels). Each item has 3 answering categories (true, more or less true, not true)
Finally we used current treatment status as an additional criterion in the evaluation. This information was obtained from the PCH professionals, who filled in a questionnaire, using both information on file and from the examination of the child during thestandard health examination, but not from the questionnaires answered by the parents.

1.2.3 IRT analyses

1.2.3.1 Item Response Theory

Item Response Theory (IRT) is a general term, describing a family of statistical models that specify how the probability of having a certain response depends on the position of a respondent (person location) and the position of a question (item location). We analyzed the data using the Partial Credit Model (PCM){Wright, 1982 606 /id}. The PCM is a unidimensional model, which means that no distinction between sub-dimensions like externalising and internalizing problems is made. The PCM is a quite simple model, more so than some other models that have many additional parameters and even multidimensional latent scales {Gardner, 2002 550 /id}. All questionnaires allow for the calculation of a single total score and therefore assume unidimensinality. However, two questionnaires also allow subscales scores to be calculated, indicating that sub dimensions are assumed. We therefore could not expect the simple PCM model to perfectly fit the data, and accepted relatively high outfit statistics up to 1.7. We used the PCM because it results in an interval measurement level, unique for more simple IRT models, and because parameter estimation is less sensitive to random fluctuations in the data. A graphical representation of the PCM is given in Figure 1.

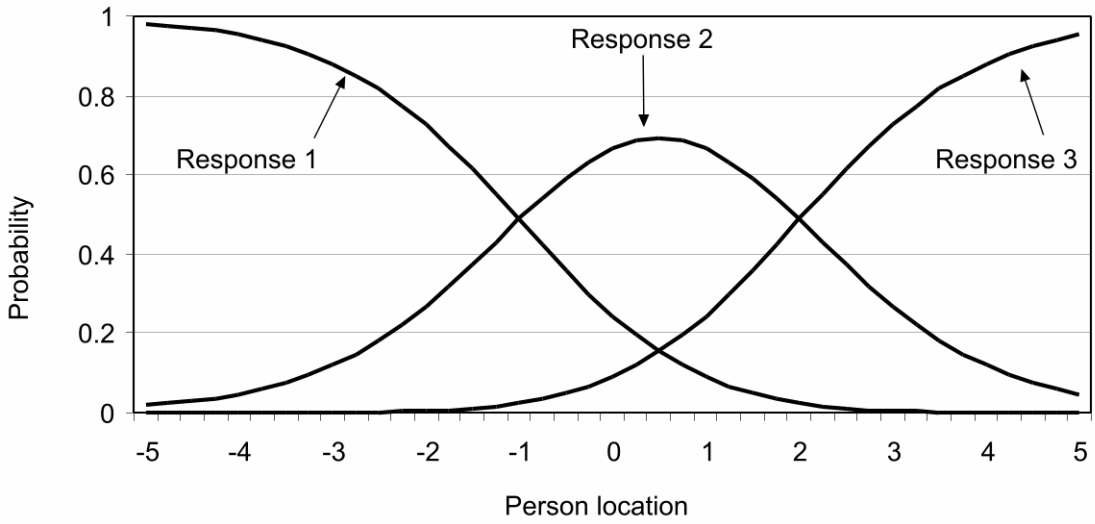


Figure 1. The probability of a response in any of the three categories

All items in the questionnaires have three response categories. Figure 1 presents the probability of the three responses to an item, as a function of the person location. A person located below -1 will most probably give response 1. When he has a location above 2, response 3 will be the most likely one and a location between -1 and 2 will most likely result in response 2. The positions on the latent scale where two adjacent categories are equally likely (here: -1 and 2) are known as thresholds. The example item in Figure 1 has an item location 0.5. This item location, the mean of the thresholds, is an indication about the point on the scale where different answers discriminate between higher or lower positions. The item location is therefore an indication of the mean level of 'difficulty', or severutry of this particular item. An item that discriminates between more severe problems will have a high item location.

We used the RUMM 2020 software (http://www.rummlab.com.au/){Andric, 2003 610 /id} to estimate item locations and thresholds in a randomly selected subset of 1650 respondents. This software can handle incomplete data matrices. The fit of items to the PCM were assessed using their fit statistics. We ignored estimation errors, as is usual in

developing CAT procedures. Items that fit the model very badly and have an 'outfit statistic' {Wright, 1982 606 /id} bigger than 1.7 - which may indicate that they actually measure something different - will be removed from the analysis. The criterion we used implies that we will accept some items that are, in Linacre's terms 'unproductive for construction of measurement construction', but no items that really damage the measurement {Linacre, 2002 609 /id}. The reason to use a somewhat less restrictive criterion is that we wanted to keep the broad scope of instruments as the CBCL.
We assessed Differential Item Functioning (DIF) in relation to gender and ethnicity in order to be certain that the estimated parameters would be valid, independently from gender and ethnicity. We did this by performing multinomial logistic regressions for all items, with the estimated persons location as the dependent variable. In the first step the raw score on the item involved was the only predictor. In the second step the interaction of the raw score with either gender or ethnicity was added as predictor. Items were considered as showing DIF when adding the interaction in the analyses led to an increase of the explained variance with more than 3.5%

1.2.3.2 Defining a cut off value for the determine a clinical score

The person location is an estimate of the amount of psychosocial problems. For the purpose of this study, this estimate needs to be classified in to a normal and an elevated range, corresponding as much as possible with the CBCL Total Problem SCAL (TPS) clinical cut-off point. We assumed this to be a valid criterion for instruments aiming at the detection of psychosocial problems. Aim of this study is to assess whether an abbreviated procedure will result in acceptable test parameters.
So, a certain cut-off point on the latent scale must be chosen that separates normal from elevated scores. A low value for the cut off point will identify many scores as problem scores and thereby increase sensitivity, but decrease specificity. A high value will do the opposite. We can choose a sensible cut off point by looking at the ROC to predict the CBCL classification from the person location estimates. We defined our cut-off point as that value that gives a specificity of minimally 0.90.

The exact estimate of a person's location will vary somewhat with the number of items used to make the estimation. The best value for the cut off point may therefore also depend on the number of items that is used to calculate the person location estimates. We suppose that in real life, large scale assessment procedures a number of about 30 items is the maximum feasible. Therefore we estimated the persons' locations in a simulated CAT with a maximum number of 30 items and performed a ROC analysis to assess the cut-off point resulting in a specificity of 0.90 and the sensitivity at that point. To check whether using different numbers of items would lead to another cut-off point, we repeated these CAT simulations with a maximum number of 20 and 10 items and also with a simulated CAT, without restrictions to the number of items, but continuing until the persons locations were estimated with 95% accuracy.

All these calculations were done in the calibration sample.

1.2.3.3 Validation of the CAT by means of a simulation study

We validated the procedure using a CAT simulation in the validation sample. In this CAT we aimed to assess whether a person scored below or above the cut off value, with a certainty of 95%. Two evaluation criteria were used: having a TPS score in the clinical range and currently being treated for psychosocial problems.

In order to assess the efficiency of the procedure, we checked the number of respondents for whom the CAT procedure reached convergence; *i. e.* for whom the procedure indeed resulted in 95 % certainty about a score above or below the cut off point. Also the number of items needed to reach convergence will be presented.

Furthermore, we calculated the sensitivity and specificity indexes at all possible cut-off points; the resulting ROC curves will be presented. To asses the validity of the chosen cut-off point, the sensitivity and specificity for the cut-off point will be presented and we will asses to what extent this cut-off point resulted in correct classification of the respondents.

1.2.4 IRT analysis

We started the IRT analysis with a total of 205 items 35 from the PSC, 25 from the SDQ, 26 from the PSYBOBA and 119 from the CBCL; two open ended questions from the CBCL were not used. In the initial analysis, 11 items had an outfit statistic bigger than 1.7 and were removed from the analysis. The IRT analysis was repeated with the remaining 194 items and another 4 items were removed because of an outfit larger than 1.7. In the third and final analysis with the remaining 190 items, the largest outfit statistic was 1.64. Only 6 items had an outfit larger than 1.50. In the remainder, we will use the item locations calculated in this final analysis[1]. The Person Separation Index, a measure of the fit of the model, was 0.93, indication a good fit.

Figure 2 presents the item locations of all items, split by questionnaire. As mentioned before, these item locations are indications of the level of severity of the items. The most extreme items on the right (measuring very serious psychosocial problems) are items from the CBCL, which appears to have more extreme items than the other three questionnaires.

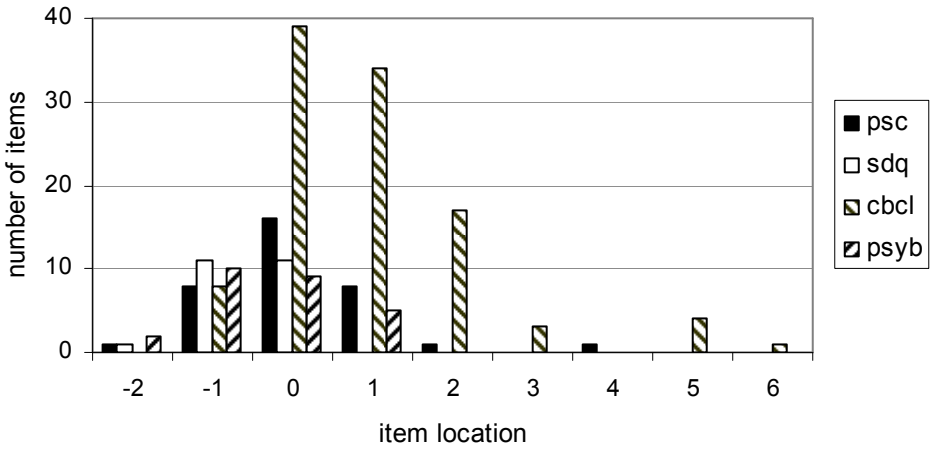


Figure 2. Item locations of the items in the four questionnaires

[1] Items removed from the analysis were: CBCL 2 Allergie (geef aan) …..; CBCL 4 Astma; CBCL 5 Gedraagt zich als een kind van het andere geslacht; CBCL 44 Nagelbijten; CBCL 55 Te dik; CBCL 56d Oogproblemen (geef aan) ……; CBCL 56e Huiduitslag of andere huidaandoeningen; CBCL 56h Andere Problemen (geef aan) ..... (remark: concerns psychosomatic problems); CBCL 77 Slaapt meer dan de meeste leeftijdgenoten overdag en/of ‘s nachts (geef aan); CBCL 97 Bedreigt andere mensen; CBCL 98 Duimzuigen of zuigen op vingers; CBCL 107 Broekplassen overdag; CBCL 108 Bedplassen; SDQ 11 Heeft minstens één goede vrien of vriendin; SDQ 22 Pikt dingen thuis, op school of op andere plaatsen;

Of these 190 items, a small number (8) showed some Differential Item Functioning. Five items sjowe DIF in relation to gender (sexual problems, running away, attacking others, being ill without physical cause and problems with teachers). Three items (tantrums, (not) being assertive, talking about suicide) showed DIF in relation to ethnicity. As most of these problems have a low prevalence and will therefore have a small overall impact on the final estimations, we decided not to remove these items.

1.2.5 Cut off value for clinical score

After the category thresholds parameters were estimated using the RUMM software, we performed a CAT simulation on the calibration sample with a maximum number of 30 items to be used.

Figure 3 presents the number of respondents by the calculated person location on the latent scale, by CBCL score divided into normal, borderline or clinical. The ROC curve indicated that with a cut-off point of -1.9 (rounded) the specified specificity of 0.90 was reached. Sensitivity at this point was 93%.

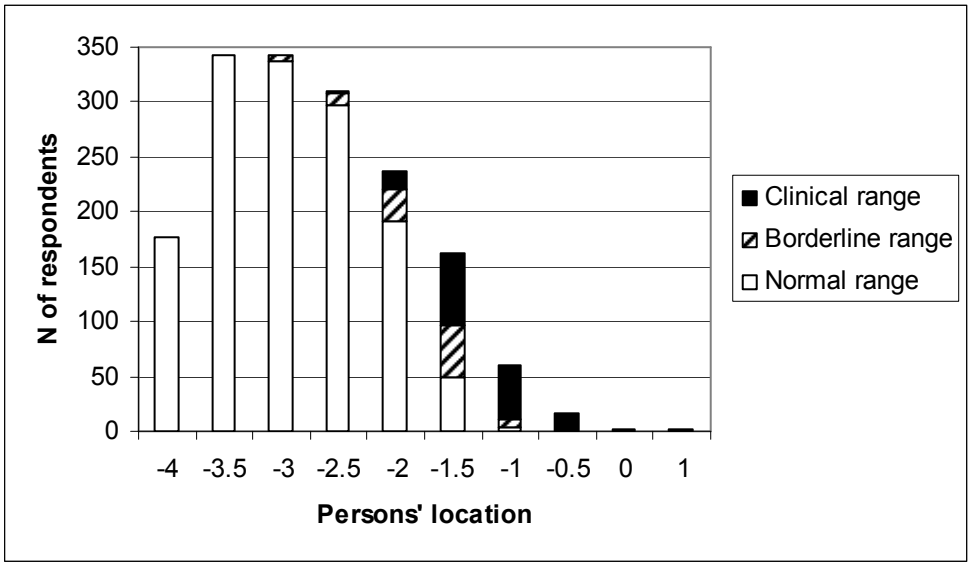


Figure 3. Distribution of estimated person locations in the calibration sample by CBCL classifications

Table 1 presents the sensitivity and the specificity at that cut off point and the AUC, when the estimates of the persons' location for CAT procedures with different number of items and a Cat with an accuracy of 95% of the estimated person's location as criterion and no limitations to the number of items to be used. The figures for the analyses with a maximum of 20 and with an unlimited number of items to be used are very similar to those from the analysis with 30 items. With fewer items to be used sensitivities decrease.

Table 1 Sensitivity and Specificity at a cut of point of -1.9 in relation to the maximum number of items used to estimate persons' locations.

Criterion	Sensitivity	Specificity	Area Under Curve
no of items used 30	93%	90%	0.97 (0.96 – 0.98)
no of items used 20	90%	89%	0.96 (0.95 – 0.97)
no of items used 10	79%	88%	0.92 (0.90 – 0.94)
5	64%	88%	0.90 (0.84 – 0.89)

estimation with 95% accuracy	92%	90%	0.97 (096 – 0.98)

1.2.6 Evaluation in the validation sample

Figure 4 presents the number of items needed in the CAT simulation in the validation sample, to reach convergence, *i.e.* to assess with 95% certainty whether the respondents had a true score below or above the chosen cut-off point of -1.9. Figure 4 presents a global overview of the results. In 40 cases (10%) convergence was not possible, with less than 100 items. They had a mean person's location of -1.88 (sd=0.18. Two of them had a clinical CBCL TPS score, 28 scored in the normal TPS range and 10 had a borderline TPS score.

For the 351 cases for whom convergence was reached with less than 100 items, the mean number of items used was 11.5 (sd: 13.0). For 47% of the respondents the procedure converged with maximally 5 items; for 66% up to 10 items were needed. For 83% 20 items were used and for 92% 30 items.

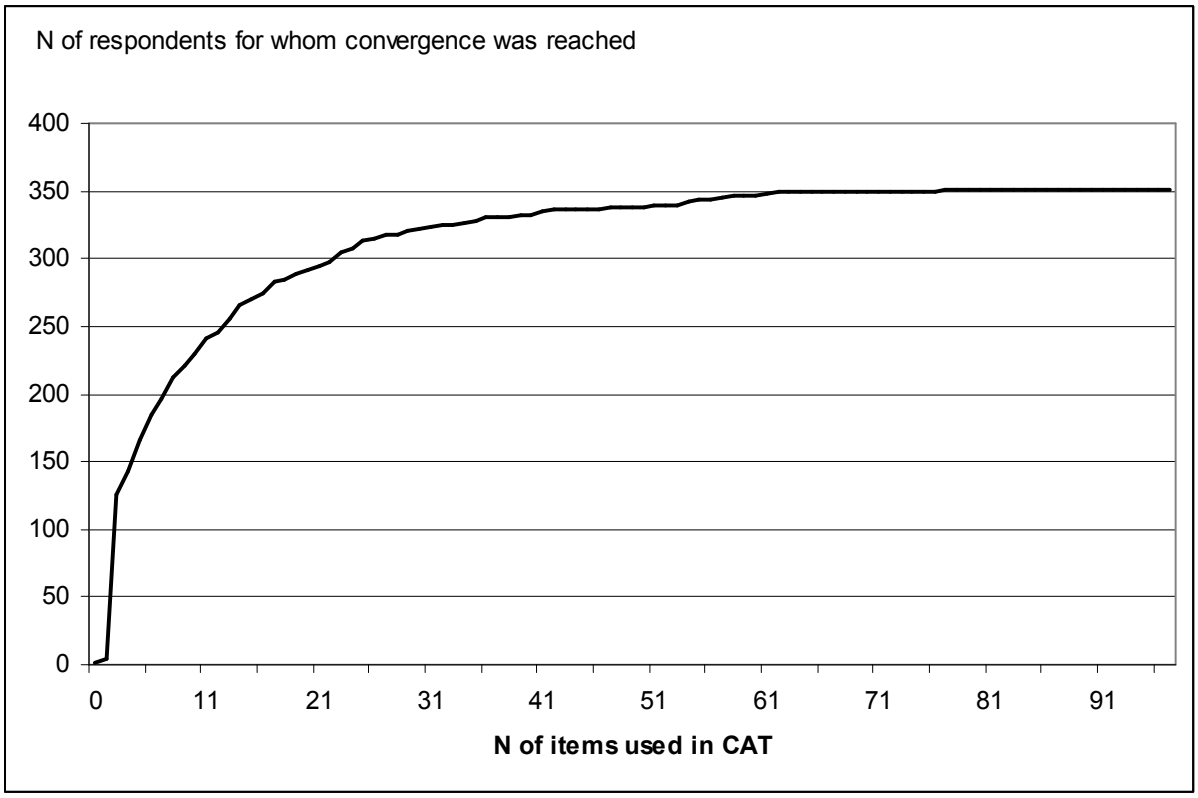


Figure 4. Efficiency of the CAT procedure: number of persons with a converged estimate by number of items needed to achieve convergence

Figure 5 and 6 present the ROC curves to predict both clinical CBCL scores and current treatment. When predicting a CBCL TPS score in the clinical range, the AUC is 0.92 (0.85 – 0.99). With treatment status as criterion it is 0.74 (0.63 – 0.84). The sensitivity for a clinical TPS score of the CAT procedure at the chosen cut-off point is 0.89 (70.6 – 97.2), with a specificity of 0.91 (87.0 – 93.3). Kappa was 0.53. The sensitivity for current treatment status is 0.55 (0.37 – 0.72), with a specificity of 0.89 (0.85 – 0.92). Kappa was 0.32.

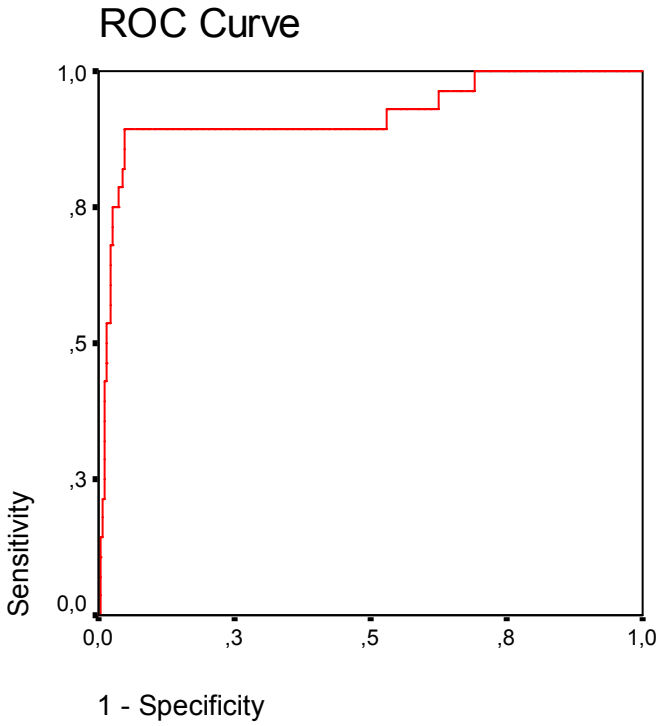


Figure 5 ROC curve for CAT simulations in the validation sample with a clinical CBCL TPS score as criterion

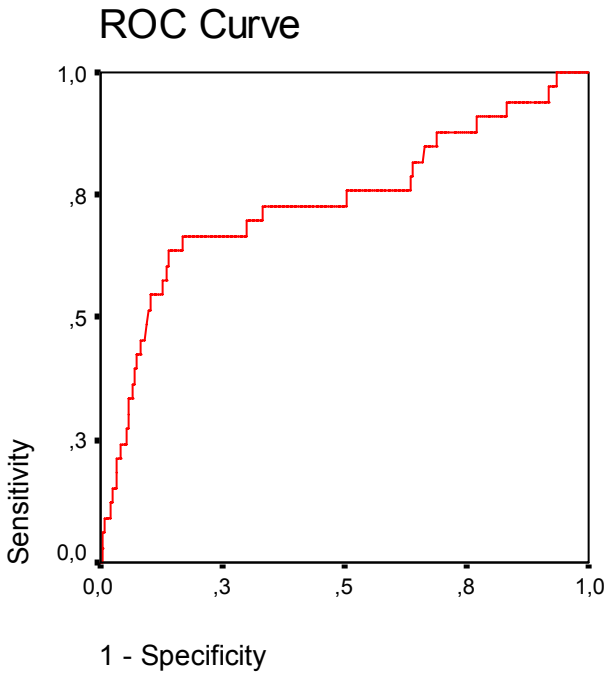


Figure 6 ROC curve for CAT simulations in the validation sample with current treatment status as criterion

Clearly predicting current treatment status is worse than predicting a clinical CBCL TPS score, probably because there is no overlap between predictor and what is predicted.

Overall, in relation to the CBCL TPS, the CAT selection procedure resulted in a correct classification of 91% of all children involved. Figure 7 presents the CAT classification in relation to having a CBCL TPS score in the normal, borderline or clinical range. The CAT resulted in a correct classification for the large majority of cases with a normal or clinical score. Most of the cases with a score in the borderline range, however, were classified by the CAT procedure above the cut-off point.

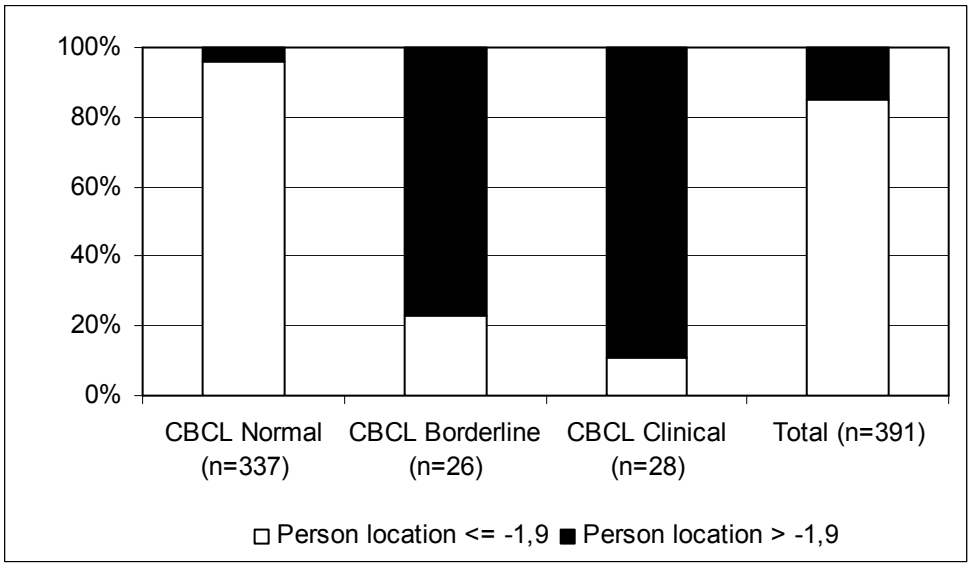


Figure 7 Classification based on theta in relation to CBCL normal, borderline or clinical scores after a CAT selection procedure.

Currently in the Netherlands the SDQ is the de facto standard used in PCH. Therefore, we also checked the relationship between the SDQ score and the persons' location estimated by the simulation. As our data set is incomplete, the SDQ data are only available 140 respondents in the validation sample. So, the results should be considered as very preliminary. Pearson's Product Moment Correlation coefficient and the estimated person's locations is 0.87. Of the twelve respondents with an elevated SDQ score, 10 also had persons' location above the defined cut off point. Only 2% of the 128 respondents with a SDQ score in the normal range had an estimated person's location above the cut-off point.

2 Developing a working prototype for a Computer Assisted Test to identify children with psychosocial problems

2.1 Introduction to the application development

2.1.1 Background and aim

Previously we showed that nearly all items of the CBCL, the PSC, the SDQ and the PSYBOBA fitted well into a unidimensional Partial Credit Model. In a simulation of a Computerized adaptive test (CAT), using answers given on a paper an pencil questionnaire, we also showed that the simulated CAT resulted in a good distinction of children with and without problems and we showed that this could be achieved efficiently, using, overall, only a few questions. This simulation study provided preliminary data on the validity of the procedure.

However, the conclusions from a simulation study like we did must be treated carefully. As we also already pointed out, there are reasons why the results from the simulation study may not be valid in a real life situation:

- The data set, on which the simulation study was based, was an incomplete data matrix; for all parents the answers of the CBCL questions were available. The other three questionnaires, however, had been answered by only one third of the sample. In the simulation part of the questions available could therefore only be used for those respondents who actually had answered them.
- In the simulation study we treated the answers on the paper and pencil questionnaire as if they were identical to answers that would be given in a CAT. This assumption may be questioned. In a paper and pencil questionnaire, respondents can easily review answers given to earlier questions, correct them or adapt answers to later questions because of answers given on earlier questions. They may, for example conclude that certain aspects of theirs child's behaviour was adequately described with items already answered and that a later item x does not really add something new. In a CAT procedure this item x may be the first one asked and may therefore be endorsed because it really says something new about the child.
- The last decade many studies have been published on using computer and/or internet based questionnaires. Many of these studies are quite positive, stressing for example that such procedures result in reliable and valid data, one of the reasons being that respondents feel more safe and are therefore inclined to give more honest answers about socially less desirable subjects. Therefore, answers to computer and/or internet based questionnaires are not necessarily identical to answers to paper and pencil questionnaires.
- Finally, there are practical reasons which may limit the usability of a CAT procedure, thereby limiting the efficiency and possibly the validity of the results. Despite the very high rate of computer literacy in the Netherlands and despite the wide spread of internet, not all parents have access to computers or the internet, not all parents are skilled and experienced computer or internet users.

Therefore, despite the fact that the results of the simulation study were very promising, ultimately, the validity, the efficiency and the practicability of an internet based CAT must be assessed in the real world, with parents answering a real CAT.

Therefore, we decided to try to develop an application, that could be used in a study to assess the validity, the efficiency and the practicability of a CAT by Preventive Child Health Care (PCH) in order to identify children with psychosocial problems.

2.2 Global description of the procedure followed

The development started with a provisional definition by the project team members of the requirements of the application to be developed. This resulted in a first document (Requirements Program, version 1, Spring 2005). This document described the requirements concerning item administration, presentation of the results to the Preventive Child Health Care (PCH) professional, operation of the application in an internet c.q. intranet environment, data management, performance, quality and security. This requirement program was meant to be used as the starting point for a Rapid Application Development model. This model implies an evolutionary development strategy, in which an application is developed in continuous interaction with relevant partners until the product satisfies them.

The Requirements Program was first discussed with four representatives from the field of PCH to assess the extent to which the requirement satisfied their requirements. This lead to some revisions of the requirements.

Before the actual application could be developed, some basic design decisions concerning the specific CAT algorithm to be used and concerning the specific information to be collected and to be presented to Preventive Child Care had to be made. These decision concerned the question which scales to assess, which item or items to use as starting items for the CAT, which criterion to use to stop assessment and which parameter to use to assess the accuracy of the (ongoing) assessment. These decisions will be described in paragraph 2.3.

Thereafter the actual aplication development started. This took about 8 months in which the developers and the TNO team members met regularly, discussing the progress of the development process, using the Program Requirements as the starting point of the discussion and adapting the Requirements Program, when necessary, due to increasing insight in the limitations of the software development tools, the consequences of choices made earlier and the usability of the application's modules when they came available.

In December 2005 a first working prototype, tested for the accuracy of the calculations became available. This prototype was presented to the representatives of the PCH, and discussed, also in relation to their remarks concerning the requirements program in spring 2005.

Their comments were noted down and classified in two categories:

- necessary changes to be implemented before the start of the test stage
- desirable changes, to be implemented before application roll out on a bigger scale (evaluation stage)

The changes labelled as necessary were implemented.

Paragraph 2.4 will describe the prototype as it is available right now.

2.3 Basic design decisions

2.3.1 Introduction

This paragraph describes the basic design decisions that were made concerning the specific CAT algorithm to be used. They concern the following questions:

- which scales should be assessed and which information should be presented to the PCH professional?
- Which item or items should be used to start the CAT?
- Which parameter should be used to choose the best next items in an ongoing CAT
- Which criterion should be used to stop an ongoing CAT of an individual child?

2.3.2 Which scales to assess and which information to present to the PCH professional

The application to be developed aims to improve the detection of children with psychosocial problems during standard health examinations by PCH. Currently, PCH uses paper and pencil questionnaires for this purpose. These are used for two – related – fucntions:

- to identify children who probably suffer from such psychosocial problems that they deserve some further attention, either a more close monitoring during some time, to see how their problems will be developing, some form of support and / or counselling or referral to some form of specialized care
- to get some insight in the kind of problems present.

The application should serve these two functions.

In the Netherlands , but also in other countries, the most widely validated indicator for likely problems is the Child Behavior Checklist (CBCL){Verhulst, 1996 536 /id}. This questionnaire was used to validate some of the questionnaires that are currently being used by PCH. So, for identifying children with problems it seemed sensible to include a Total Problem Score, comparable to the CBCL Total Problem Scale in the assessment.

The CAT score is calculated according to the Partial Credit Model{Wright, 1982 606 /id}, and has the advantaged that it results in a measurement at an interval level. That means that a difference of two points on that scale can be considered as being twice as large as a difference of 1 point. Yet, this scoring is little more than locating children on some position on a quantitative scale. This position in it self gives not much information about whether further attention is desirable. The CBCL TPS allows for the classification of children in a normal, borderline or clinical range and this classification has been shown to be clinically relevant. In the first part of this report we have shown that we could reproduce this CBCL classification rather accurately and we therefore decided to present that information to the PCH professional, too.

PCH uses paper and pencil questionnaires also to get some insight in the kind of problems present in the child. He or she does so, using subscales if they can be calculated and by inspecting the answers on individual items. We decided that the CAT should offer similar functionality. So, the application should offer the possibility to

inspect the answers on individual items and scores on some relevant subscales. The question to be answered, then was, which subscales to include.

One of the most common distinctions is that between internalizing and externalizing behaviour. The first category refers to problems like, depressive feelings and withdrawn behaviour. The second one to aggressive, anti-social behaviour, conduct disorders and the like. The CBCL contains two (broadband) scales referring to these very categories. Explorative principal component analyses on the PSC, the PSYBOBA and the SDQ suggested similar components in these questionnaires {Vogels, 2005 561 /id}. The distinction between internalizing and externalizing behaviour also has the advantage that it is widely known by health care providers. We therefore decided to try to assess also the extent of internalizing and externalising problems present. Finally we decided to try to assess the amount of hyperactive behaviour with the prototype to be developed. It is a far more specific category of problems than internalizing and externalizing, but receives a lot of attention currently. The questionnaires contain far less items on hyperactivity than on externalizing and internalizing and we thought it would be worthwhile to see if we could achieve some indication of hyperactivity nevertheless.

Three (CBCL, PSC and PSYBOBA) out of the four questionnaires do not measure the concept of hyperactivity as such. Similarly, three questionnaires don't measure Internalizing and Externalizing as such (SDQ, PSC and PSYBOBA). So we asked one paediatrician, one public health medical doctor and two psychologists to classify all items in the questionnaires as to whether they could be considered indicators of that type of problems. Discrepancies between the classifications were dissolved in discussion between the two psychologists.

Just as we had done for the total sample, we repeated the Item Response Theory Analyses in the calibration sample using only the Externalizing, Internalizing or Hyperactive items, in order to determine item locations. As expected, the item locations correlated almost perfectly with the item location on the CAT's Total problem scale. We also determined the cut-off point for these subscales at those points that resulted into a specificity of about 0.90. This wsa the same point (-1.9) as we had chosen for the Total Problem Scale. Sensitivities at those points were 0.85 for Internalizing, 0.92 for Externalizing (with the CBCL's syndrome scales as criterion) and 0.96 for Hyperactive Behaviour (with the SDQ's Hyperactivity scale as criterion.).

2.3.3 Choosing items to start the CAT with

There are several options to consider when deciding which item to use as the first item in CAT measurement. The first option is to use an estimate of the expected person's location and to choose that item that would give the most extra information. That would result, theoretically, in the most efficient measurement process. We have little information that could guide such a first estimate. We could use the best estimate available, i.e. the mean score of the population and select that single item most closely located to the population mean. The literature {Wainer, 1990 604 /id} advises against this option, as it could lead to the population getting used to this item. A further argument against this option is that the population mean is not a very accurate estimate of the individual person's location. Using this option would therefore not lead to a big increase in efficiency.

Another option would be to use a randomly chosen item as the first one. This has the advantage that each administration would be different, also in repeated measurement on

one individual and that there will be no customisation to a constant first item. A second advantage is that it would be an unbiased starting point and that the starting item would have no systematic effect on the final measurement. This option has also disadvantages: it could lead to a less efficient measurement process and the first item, chosen randomly, can be a very extreme item, with an item location on the top of the scale. This could lead to parents being shocked.

When the criterion to stop the measurement (see below) would be certainty that the person involved scores above or below a predefined cut-off point, one could consider using an item near the cut-off point. In that case, it would lead to an efficient measurement procedure, but possibly also to customization in the population. Also, this option would lead to many parents being confronted with extreme, possibly deterring items.

A final option, and the option we decided to use, is to start with a subset of items, varying in content and predictive of the final estimate. This option has several advantages: respondents can make mistakes when answering questions. With a single starting item, a single mistake can have a substantial effect on the final results, when convergence is reached very soon. When using more items, a mistake with the starting item will decrease the accuracy of the estimates made in the beginning of the assessment, due to inconsistent answers and convergence will there fore take longer. But the effect of such mistake on the final estimate will be minimal. Using a subset of items also offers – in our case – the possibility of choosing items from the different subscales to be assessed. This makes certain that the broad scope of questionnaires as the CBCL and the SDQ will also be reflected in our assessment. We decided not to use a random sub set, as this would imply that extreme items could be chosen as starting items. Instead we chose for items that were endorsed by a relatively large part of the respondents that had good predictive power for the final estimates and that were distributed over the three subscales Internalizing, Externalizing and Hyperactive Behavior. We used two items from the Internalising and Externalising scales and one for the Hyperactivity scale.

The items used as starting items were:

CBCL 3: Spreekt veel tegen of maakt veel ruzie
CBCL 10: Kan niet stil zitten, onrustig of overactief
CBCL26: Lijkt zich niet schuldig te voelen na zich misdragen te hebben
CBCL 32: Vindt dat hij zij perfect moet zijn
CBCL 88: Mokken, pruilen

2.3.4 Choosing next items

In order to determine which next item should be chosen as the next one to be administered, we decided to stick with the most widely used criterion: Fisher's Information Criterion. This criterion works fast and other options, like Potential Impact and Attraction Measure) do not lead to better results, {Vestering, 2004 605 /id}.

2.3.5 When to stop the assessment of the Total Problem Scale score?

Traditionally, in IRT based CATs, the assessment continues until the person's location has been estimated with a certain predefined accuracy. This is, in principle, a very elegant procedure. This procedure is not very attractive in our case. The reason for this is that there is a discrepancy between the distribution of the item locations and the

distribution of the persons' locations, as is shown in figure 8. Most of the persons' locations are on the left of the scale, centred somewhere around -2.5, indicating few problems in most respondents. Most of the items are located in the middle of the scale, centred around 0, indication that they best distinguish between persons with more problems.

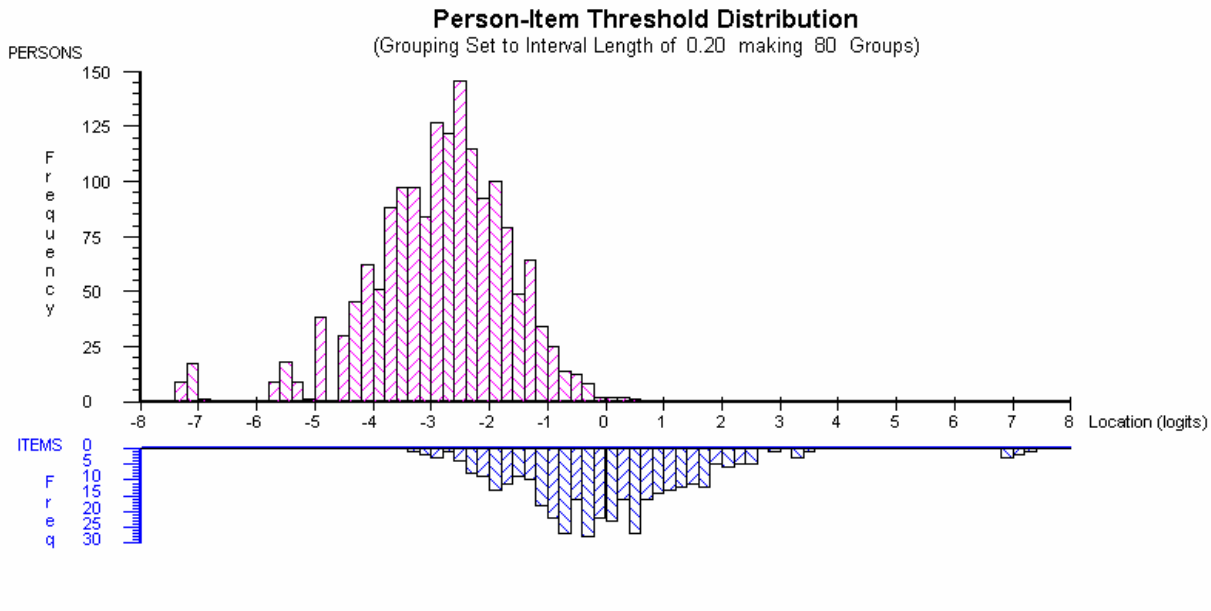


Figure 8. The distribution of item and person locations

As a consequence, the CAT would have difficulty to achieve the predefined level of accuracy for many of the respondents with relatively few problems. In any case, it would need many items for those respondents. And the majority of the respondents have few problems.

The aim of the application, however, is not a full assessment of each individual child, but to identify those children that probably need further attention, because of a relatively high score on a total problem scale. We therefore decided that we would not use the usual CAT criterion of a measurement with a predefined accuracy overall. In the first part of the report we showed that using a specific cut-off point, the CAT enabled a sensitive en specific distinction between children with a clinical CBCL Total Problem score and children with a lower score. We decided to use this as the criterion to stop the CAT. In other words, the CAT continues until the procedure results in an estimated person location with a 95% confidence interval that does not include the cut-off point of -1.9.

This criterion is in concordance with the task of Preventive Child Health Care to identify children with problems and ensures that parents with children with few problems do not have to answer many (mostly irrelevant) questions. Of course, it will not always be possible to conclude with 95% confidence that a child scores above or below the cut-off point, either because of a score too close to the cut off point, or because of – seemingly – inconsistent answers, resulting in a large confidence interval. Such children are to be considered as a kind of borderline cases, needing further assessment during the following health examination.

Figure 9 shows the number of items needed when using the usual stop criterion and the criterion we decided to use, in relation to the CBCL TPS (normal, borderline or clinical range). The figures are based on the validation sample (n=391) and include all children, also those for whom the CAT did not converge. It shows that using our criterion results in a far more efficient assessment for children with a CBCL TPS in the normal range. For children with a TPS in the borderline range (n=26/391) our procedure is less efficient. This is to be expected as our procedure will try longer to assess whether the child scores below or above the cut off point. As a matter of fact, the procedure does try it probably unnecessarily long, as these children, according to the CBCL classification fall in the borderline category, indicating that the classification is uncertain.

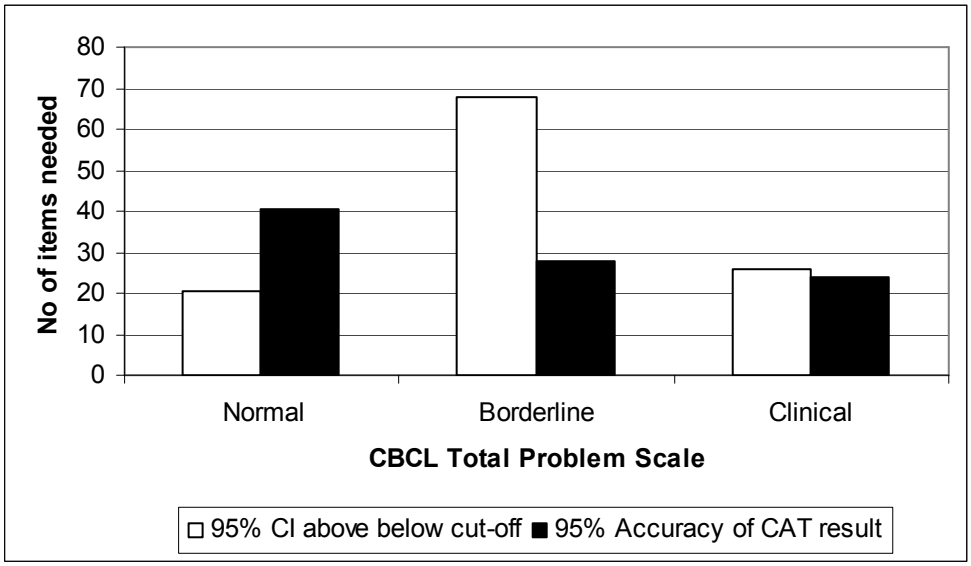


Figure 9. Mean no of items used in two types of CAT: a) estimating whether the 95% CI of a persons location lies above or below -1.9 and b) estimating the persons location with an accuracy of 95%

Despite the efficiency of our procedure, the number of items presented to a part of the respondents is too high, especially as we wanted not only to measure the Total Problem Scale score, but also scores in Internalizing, Externalizing and Hyperactivity.

We therefore decided to use an additional criterion to stop the assessment, namely a maximum number of items to be administered. Assuming a number of about 30 as the maximum feasible and four scales to be assessed, we decided to use a maximum of 15 items for the initial assessment of the Total Problem Scale. This would lead to a converging CAT for 69% of the respondents.

After the initial assessment of the Total Problem Scale score we decided to use a maximum of 8 items (including those items used in the start of the assessment) for the assessment of scores on Internalizing, Externalizing and Hyperactivity. After the assessment of the subscales the information collected for the subscales would be used for a final assessment of the Total Problem Scale. We assume that this will lead to convergence for more persons than the 69% achieved in the final assessment, although in some cases – with isolated problems, without much comorbidity – the information on the subscales may be inconsistent with the first assessment, resulting in a larger confidence interval of the estimated person's location.

2.3.6 When to stop the assessment of Internalizing, Externalizing and Hyperactivity?

The stop criteria for the subscales are based on the same consideration as the stop criterion for the Total Problem Scale. However, in our view the assessment of the Total Problem Scale is more essential for the identification of children with problems than the assessment of the subscales, which give additional information on problems present in the child. We therefore decided to be satisfied with somewhat less accuracy here. We will use an accuracy of 90% here and – as already described above – we will use a lower maximum number of items to be administered to parents.

2.3.7 Summary of the CAT procedure developed.

In summary, the CAT procedure as developed for the CAT application is as follows:

1 The procedure starts with asking each respondent a fixed subset of 5 questions; 2 referring to Externalizing, 2 to Internalizing en 1 to Hyperactivity
2 The assessment of the Total problem Scale score continues until we know with 95% accuracy whether the score lies above or below the cut off point of -1.9 or until 15 items have been administered; each following items is chosen using Fisher's Information Index in such a way that is maximally improves the estimation of the person's location
3 For Internalizing, Externalizing and Hyperactivity successively the procedure in step 2 is followed, but with an accuracy of 90% and a maximum number of items to be administered of 8, including those used in step 1.
4 After step 3 the person's location on the Total Problem Scale is reassessed, using the extra information collected in step 3.

2.4 Global description of the application's functions

2.4.1 Introduction

This chapter will describe the main functions of the application. The description is based on the final version of the Requirements Program, which is attached to this report as Appendix A.

The first version of this program was written by the first author, in discussion with other TNO-researchers with a background in PCH. This version was discussed with three PCH professionals. Their remarks lead to adaptations and this version was the starting point of the actual development process. This process is described in paragraph 2.5. In this stage regular meetings were held between the researchers and the developers, to assess whether the development still followed the specified requirements and whether adaptations in the requirements were necessary.

When the application was ready, it was demonstrated to and discussed with four PCH professionals. Their remarks and wishes, again, led to some adaptations of the Requirements Program. These adaptations were distinguished in two categories:

- adaptations to be applied before a pilot implementations could be started
- adaptations to be applied after a pilot implementation but before large scale distribution.

Adaptations in the first category were applied and led to the current available version of the application.

2.4.2 General

An essential characteristic of the application is that it is developed as a generic application, allowing for the administration of different investigations, *i.e.* one or more questionnaires in combination with a specific target group of respondents.

The Requirements Program distinguishes three types of users of the program:

1 Administrators: the **central administrator** determines which questionnaires will be made available in the application; the **local administrator** is responsible for managing user accounts and the definition of investigations, defined as the (combination of) questionnaires to be administered to specific target groups.
2 Preventive Child Health Care (PCH) professionals, physicians and nurses having access to the results on the questionnaires for children and there parent under their care.
3 Clients: respondents (parents, children, teachers) who have been invited by PCH to participate in the assessment.

In the current prototype all administrator tasks, central and local, are carried out by TNO Quality of Life. When the application is implemented as the initial production version, TNO will continue with its role as central administrator. However, the role of local administrator will be fulfilled by PCH.

The application is available via the Internet, reachable through separate web addresses for clients, answering questionnaires and PCH professionals, inspecting the results. Each function of the system has its own user interface, requiring separate logins for access.

2.4.3 Administration of questionnaires to clients

Clients are invited by their PCH professional to participate in the internet assessment and are given a login name and password. When they log in, they are presented with a screen, with a short instruction, stressing that they are free to cooperate and that only PCH professionals will know what they report in the assessment.
After reading the instructions (Figure 10), they are presented with the questions, one question at a time. The clients can answer a question by first selecting a response and then pressing the next-button (labeled “Volgende”) using their mouse. They can also hit the enter-key after they have selected an answer, since the next-button automatically gets focus.

Figure 10. The instruction page of the Client Application

Clients are able to interrupt the assessment at any moment they would like to do so. They can always decide to continue the assessment some other time by logging in using their old username and password. In case clients continue with an unfinished assessment and log in into the application, they will be presented with the beginning of the not fulle completed questionnaire. The system will terminate the assessment after all the questionnaires have been completed. The assessment will also stop if there is no screen activity for 10 minutes.

We have decided not to offer direct feedback to the clients about the results of the assessment, as the information could easily be misinterpreted. Instead, after completion of the questionnaires, clients are notified on screen that their PCH professional will discuss the results with them during their next consultation.

Each assessment consists of a set of questionnaires, pre-defined by the local administrator. This set defines the order in which the questionnaires are presented. The application allows for other sets of questionnaires to be administered, if and when such questionnaire sets are made available by the central administrator.

The authentication of a client is based on a username and password. This username is non- discriminatory, meaning that it contains no personal information of the client or child. After a successful login, the system looks up the corresponding session based on the username. Since each session contains a unique username, only one session can be linked to the client. In other words, if the same parent is asked to participate in two assessments, he will be given two separate usernames.

In the current prototype the only assessment available consists of three questionnaires, namely two non-IRT and one IRT questionnaire. The initial questionnaire is non-IRT, used for collecting some personal information about the child. It is also used to check if the child's age is in the required age range (set between 4-14 years). If this validity control fails, then the system displays a message informing the client that the session will terminate. Otherwise, the client can continue with the session. In the latter case, the next set of questions is fetched from an IRT driven CAT questionnaire on emotional and behavioural problems. This set of questions consists of two parts; a fixed set of five pre-selected questions with which the procedure is started, followed by a variable number of questions, depending on the answers given by the client. Figure 11 displays a question taken from the IRT questionnaire. The final questionnaire is non-IRT, used for evaluating this way of collecting information.

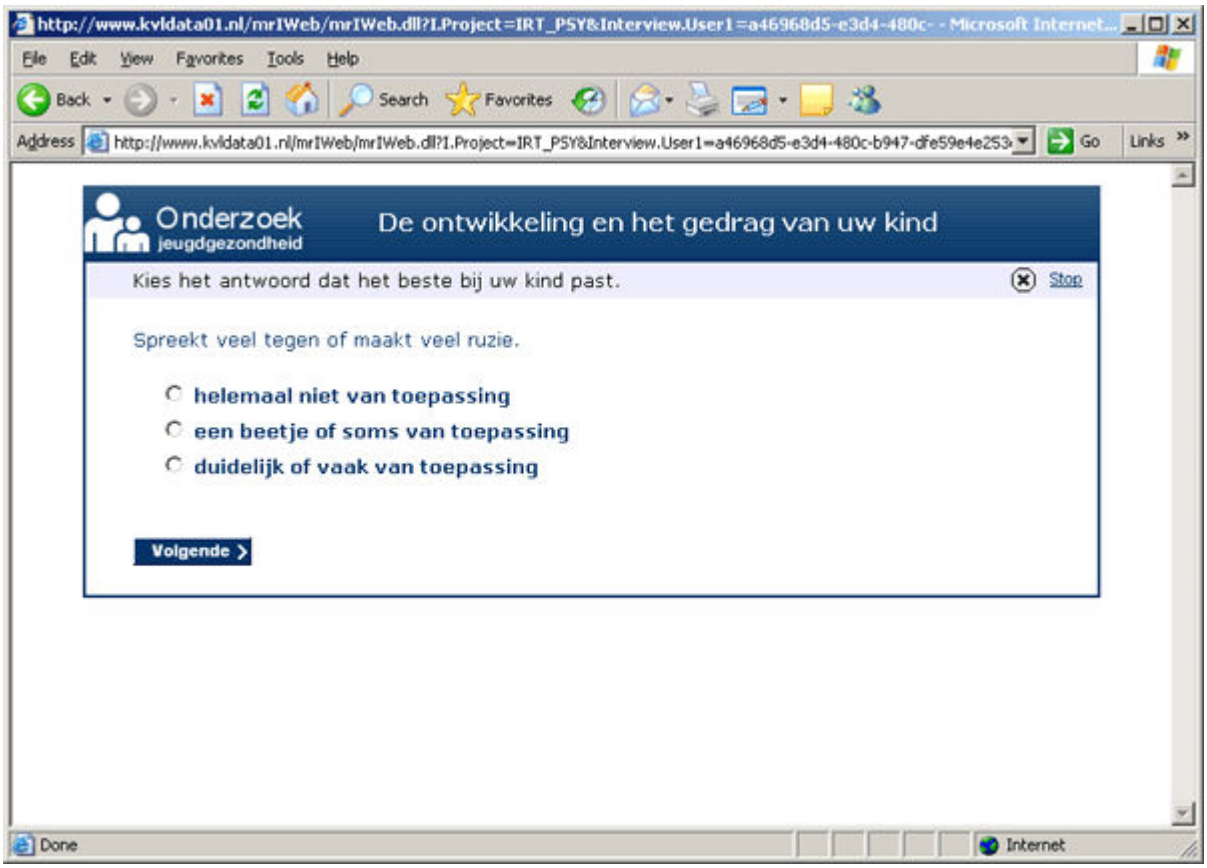


Figure 11. A example of question with response alternatives taken from the IRT questionnaire

2.4.4 *Feedback of the assessment's results to PCH professionals*

Access to the results of assessments is limited to authorized PCH professionals. They can log into the system using their username and password, provided by the administrator. After logging in, a list of available investigations (combination of questionnaires and target group) is presented, from which they can choose. Upon choosing for a specific investigation, a list of persons for whom results are available is presented (see Figure 12). Access is restricted to assessments of clients under his or her care. In other words, a PCH professional can not view the result of an assessment by a client not connected to him.

Using either a client's identification code (in the future BSN, the Citizen's Service Number), name, and / or birth date, the PCH professional can search for specific clients whose results he or she wants to see.

Persons for whom results are available are marked with "Ja" on the right of their row. Clicking on "Ja" leads to the first results screen, which is a graphical representation of the scale sores. Scales are presented as bars, with colours indicating the normal (green), borderline (orange) and clinical range (red) (Figure 13). Scales scores are represented by the arrowhead above the scale pointing to a specific position on the scale; the line around the arrow indicates the 95% confidence interval.

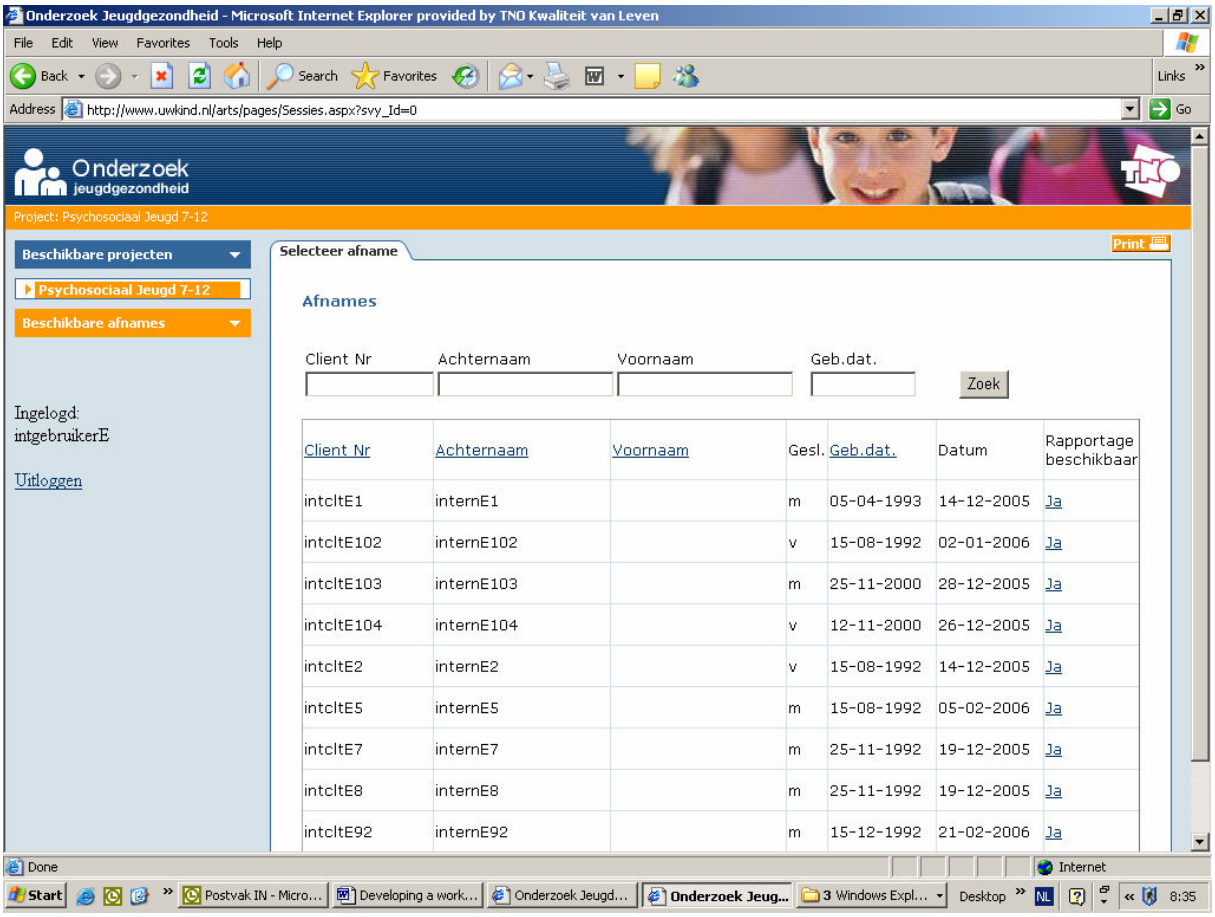


Figure 12. Listing the persons for whom results are available

Clicking on the tab 'Items' leads to the second results page, which lists the specific questions that have been asked in this assessment, the answers given and an indication of the seriousness of these answers (as in Figure 14). The list can be sorted on seriousness of the answer given and on scale, allowing a quick overview.

Results can be printed, to allow filing it in a paper register. Provisions have been made, as stated in the requirements program, to enable exporting results to an electronic register. This function has not yet been fully implemented, for there is no standard to communicate between different types of electronic registers in the Dutch PCH.

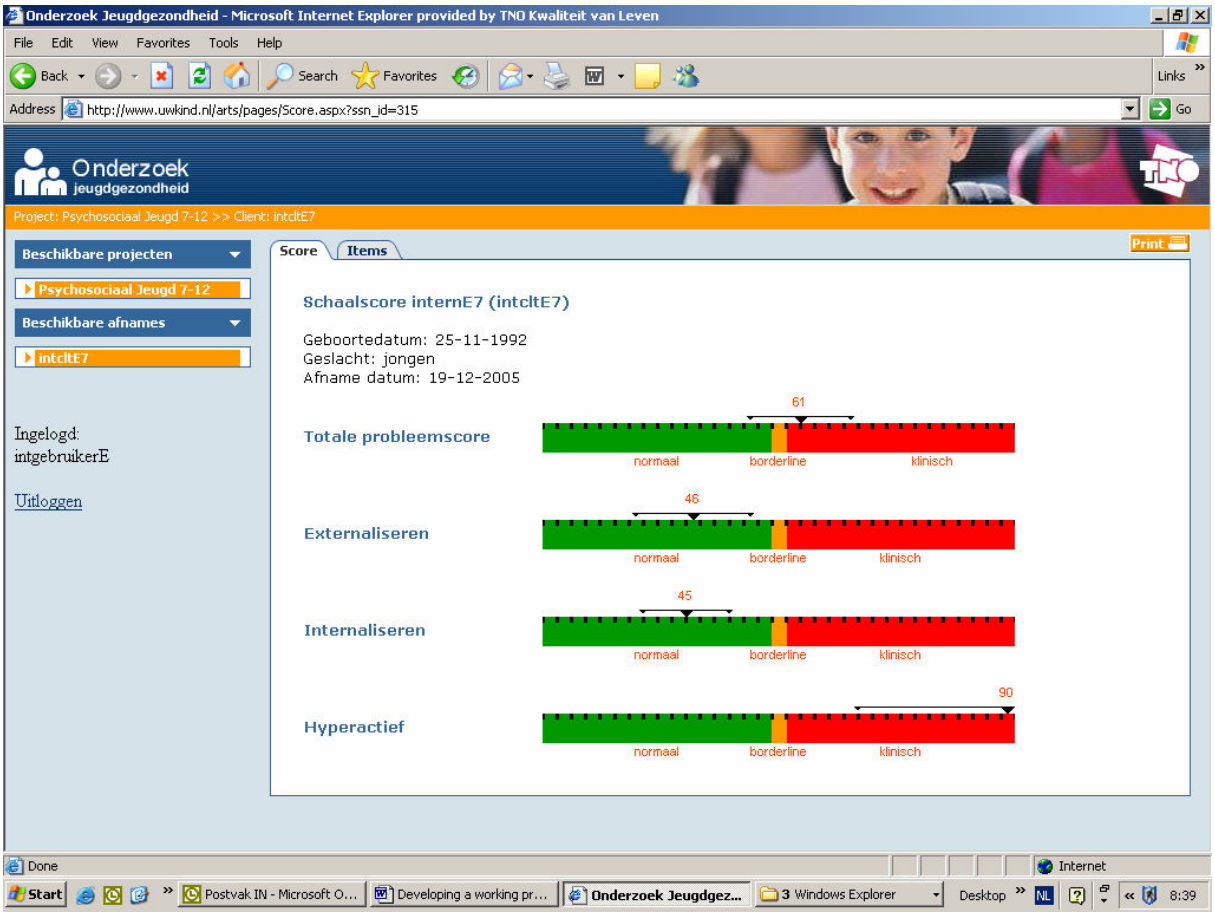


Figure 13. The first results screen: a graphical presentation of the scale scores

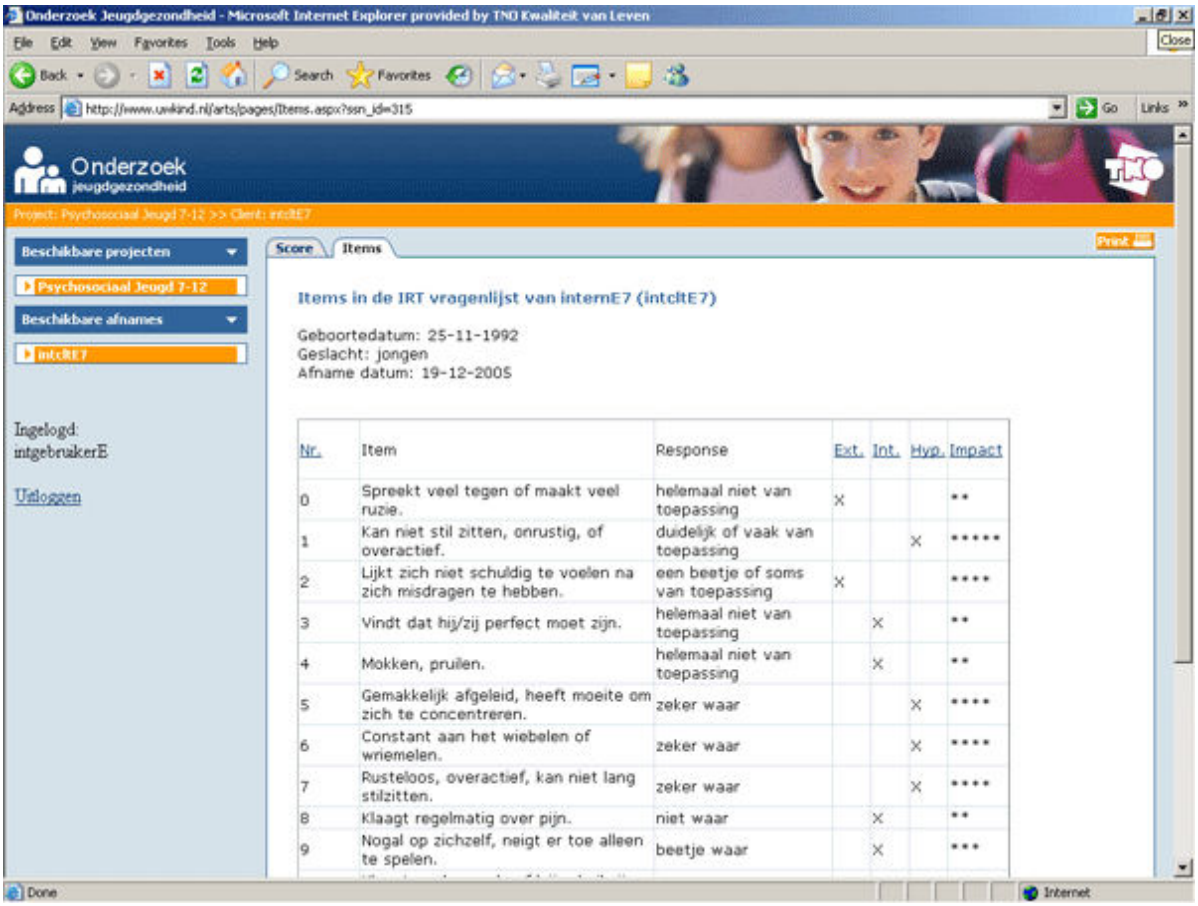


Figure 14. Second Results screen, presenting questions asked and answers given

2.4.5 *Other properties of the application*

The application is secured against identity theft and unauthorized access, based on a subset of the NEN 7510 standard for information security. The prototype satisfies the requirements found in the Requirements Program under “Security Requirements”, Category VII of the section Non-functional Requirements.

All functions of the system are protected by a system of usernames and passwords. During the test stage, this information will be provided by the central administrator. As the system is Internet based, we have decided that no data will be asked that would permit the identification of individual persons without knowing the combination of username and the person’s identity. These data have to be managed in a person register by the PCH and or local administrator using the application.

If requested by respondents the central administrator can delete individual records, as required by law. Results can be exported to SPSS by the administrator.

2.4.6 *Requirements to be fulfilled after the test stage*

The application as it stands now is a working prototype, to be used in an evaluation study. Some requirements have been foreseen, but have not yet been implemented:

1 All future functions will be available to authorized persons only (exporting data by local administrators, managing the persons register (by local administrator or local

staff), data management, definition of investigation (by local administrator or PCH professional).
2 Access rights of PCH professionals have to be defined on a local level.
3 Using the application to administer questionnaires in a school environment (for teachers and or children) should be enabled, as soon as questionnaires suited for such use, are available.
4 The application should allow for reporting result on an aggregated level, e.g. on class, school or local level).
5 Matching login names and identification codes in the PCH administration must be automatically enabled – for authorized users
6 The application should enable for a report about changes over time when more than one assessment of the same person is available.
7 Improving the application security by enforcing the NEN 7510 standard for information security.

2.5 Technical description of the application development

2.5.1 Approach

The application development took places in different stages, as illustrated in Figure 15.

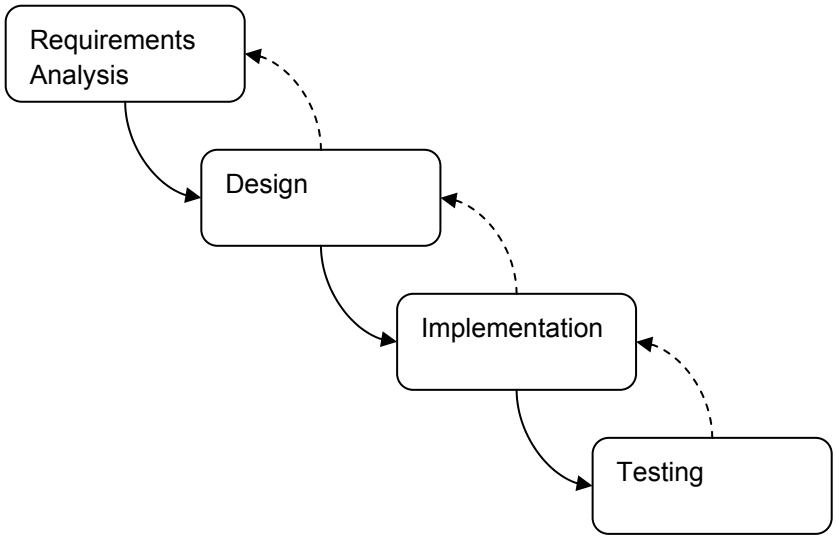


Figure 15. Stages in the development process

The Requirements Analysis gives an Analysis Model of the system with all its necessary components, describing the application domain. This model is based on the system specification, extracted from the requirements. In the development stage, this system model was transformed into system design, during which the system was decomposed into smaller subsystems. These subsystems were then implemented as independent modules, using various techniques and programming environments. Strategies were selected to build the system, making sure the modules were implemented efficiently. These strategies included making decisions about the required hardware and software platform, data management, and access control. The completion of the development strategies of the components was followed by the actual implementation.

Testing was done continuously during and after implementation of every component. As mentioned before, subsequent versions of the modules were also presented during group meetings, to elicit criticism and suggestions for improvements. The testing of each component took place in a test environment. Only when the test results were satisfactory and when the components worked as a whole, the components were installed in the prototype environment. This same environment will be the future production environment for the production version.

After extensive discussions about the requirements with the whole project team, it became clear that the following system components were needed to develop a system based on the requirements program:

- Questionnaire component (consisting of a set of questionnaires)
  - IRT – CAT driven questionnaires based on a CAT Algorithm
  - standard 'linear' questionaires
- Client Application
- PCH professional Application
- Application Database

The resulting system model including the association between the components is depicted in figure 16.

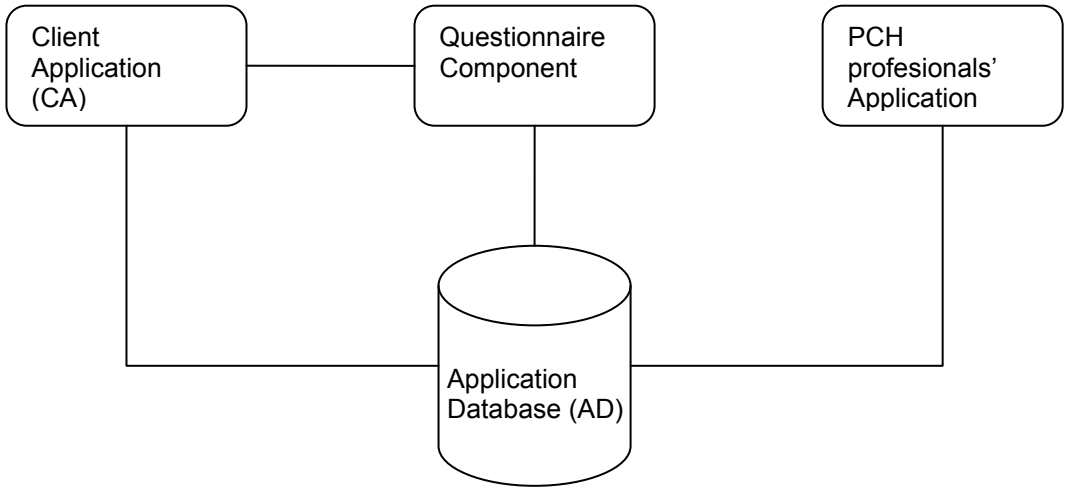


Figure 16. Analysis Model with system components

2.5.2 System description

The Client Application (CA) retrieves login data from the Application Database (AD) for user authentication. After a successful login, the CA acquires data from the AD regarding which questionnaire to start. The CA receives the web address of the questionnaire and forwards the client to the appropriate questionnaire. The initialized questionnaire first verifies the client using login name and password and data from the AD. After successful verification, the administration of the questionnaire is started. After completion of a questionnaire, the questionnaire id is stored in the AD, after which the client is forwarded back to the CA. The CA asks the web address of the next questionnaire to be started until all questionnaires have been completed. The CA displays a notification after completion of the final questionnaire in the specific investigation.

A PCH can look up assessment results through the PCH professional Application (PA). The PA retrieves login data from the Application Database (AD) for user

authentication. After a successful login, the PA displays the investigations associated with the PCH professional. Upon choosing for a specific investigation, the PCH professional is presented with a list of available assessments of individual children. Clicking one one of them results in the graphical presentation of scale scores. A second tab presents the questions asked and answers given.

2.5.3 Questionnaire Component

One of the main functions of the application is the administration of questionnaires to respondents. For this to be possible, questionnaires have to be available online. This enables clients to answer questionnaires on the Internet. The results of the questionnaire assessments have to be stored, in order to be available for the PCH professional. Therefore, a tool for building questionnaires in a web environment and one with adequate data management capacity, was needed.

The first option of software to consider was MR Interview (MRI).

To make sure that applicable questionnaires could be built with MR Interview, it was tested on the following criteria:

- Web based questionnaire (customized questions with responses)
- Dynamic assessment (for IRT feature)
- Storage and management of results
- Transition time (time after providing an answer and waiting for the next question to appear) of max. one second between questions

During this evaluation a new version of MR Interview became available (version 3.0). We decided to use this new version, since it was a completely new product compared to the previous version and contains functionality not found in previous versions and needed for the IRT application.

MR Interview (MRI) is used worldwide for building and administering web based questionnaires. An MRI questionnaire consists of two parts, namely Meta data and Routing. The Meta data is used for the static part of a questionnaire. It contains the labels of the questions and the respective responses. It also includes the labels to any additional supporting text, such as banners and notifications. The Routing section is the dynamic part of a questionnaire, containing the questionnaire logic, i.e. the order in which the questions should be presented and the action to be taken given a particular response. The routing can be customized using a scripting language (MR Script), to define the behaviour of the assessment. During an assessment, questions are presented by the system and responses are selected by users. If the results of an assessment are to be displayed for later use, some kind of data management is needed to store and retrieve assessment data. MRI provides this feature through the MR Data model, a data layer based on a SQL Database. SQL queries can be executed to retrieve assessment data.

During the feasibility study, a questionnaire containing only a few questions was used. The transition time was far below the specified transition time of one second, but it was unknown if this value would hold when more questions and and a complex routing system, needed for the IRT driven CAT, would be used.

The system model depicts the questionnaire component as a single component for ease of understanding. In reality, the questionnaire component is a collection of separate

questionnaires. The requirements specify the following questionnaires as needed for the prototype:

- IRT questionnaire (IRT), including the CAT algorithm
- Introduction questionnaire (linear; non-IRT)
- Evaluation questionnaire (linear; non-IRT)

A (simplified) model of the questionnaire component is shown in Figure 17.

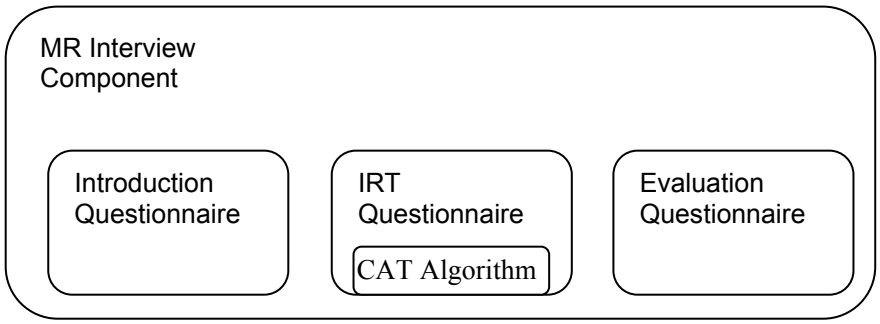


Figure 17. MR Interview Component with static and IRT questionnaires

2.5.3.1 Implementation

The next step was the actual implementation of the questionnaires. The first questionnaire to be implemented was the IRT questionnaire. The Meta data section of this questionnaire had to contain the questions from the selected item bank. The item bank was a text file containing the necessary questions, three response categories per item and the IRT parameters needed for the CAT procedure. Since the item bank contained 190 items, a manual development of the IRT Meta data section would have required a substantial amount of time. An efficient solution was to generate the Meta data section.

First, the items from the item bank were migrated to a temporary SQL database. The SQL Data Transformation Services (DTS) was used to copy data from the text file to the temporary database. A simple script generator was developed in C#, reading the items from the database and creating a string resembling the label of an item in de Meta data section of an MRI questionnaire. For every item in the database, a string was generated and written to a text file. This way of generating the Meta text file takes no more than one second. The content of the generated text file was copied and then pasted in a blank Meta data file, creating the necessary IRT Meta data section.

The Routing section of the IRT questionnaire was implemented next. Its function was to present the questions to the clients, retrieve the selected answer, use the IRT algorithm to calculate a new estimate of the persons' location, using the new answer and to decide which item to use next. Since MRI 3.0 contains a scripting language, it was decided to use this language to implement the algorithm. The algorithm was based on a test version implemented in JavaScript, which was developed during the IRT analysis stage. Although JavaScript code is in some way similar to MR Script, some differences in the MR Script language prevented a straightforward conversion of the algorithm to MR Script

2.5.3.2 Simulation

Each of the functions was tested after completion by simulating an assessment using the JavaScript application as criterion. The calculated values of the function in MR Script were then compared to an identical session of the JavaScript version.

2.5.3.3 Results

After completion, the IRT Questionnaire was presented during a project meeting. Because of the extensive use of parameter passing, a lot of processor resources were consumed. This resulted in a transition time of up to six seconds, where a maximum of 1 second was required. Another solution was therefore needed to bring the transition time down to at most one second.

2.5.3.4 DLL Component

A new approach for the implementation of the IRT-algrorithm was therefore developed: re-develop the algorithm, compile it as a DLL, and use it as an instance of a COM Object to improve performance.

The algorithm was re-written in C#, using an object orientated approach. Instead of one main class file, the functionality of the algorithm was split accordingly into separate classes. This resulted in clear and clean code, compared to the one class approach in the MR Script version.

The C# version was, also, tested by simulating an assessment and comparing the values of the estimated persons' location and the retrieved questions with a parallel simulation using the JavaScript application. Once the C# version resulted in the same results as those from the JavaScript version, the C# class files were compiled as a registered DLL.

This DLL is used in the MR Script as an instance of a COM Object. In the new situation, the MR Script in the Routing section retrieves a client response and passes the value as a parameter to the COM object. The response is the id of the next question to be asked. The script retrieves the question associated with the retrieved id and displays it to the client.

As we had hoped, the transition time was reduced dramatically using the DLL solution, down from six seconds to an almost negligible 0.2 seconds.

2.5.3.5 Linear questionnanires

The introduction and evaluation questionnaires were implemented without using any complex routing algorithm. These questionnaires are linear, i.e. every assessment results in the same fixed set of questions. The development of both questionnaires was straightforward, using standard MR Interview provisions.

2.5.4 Client Application

Clients that wish to participate in the online assessment are given a username and password for authentication and a URL or web address of the interface where they can login to start the administration of questionnaires. The application that verifies user authenticity, and redirects the user to the different questionnaires is the Client Application (CA). In other words, the CA manages the questionnaires to display to the client.

After a successful login, the CA retrieves the unique assessment identification (session id). This assessment contains information regarding the user, the status of the assessment, the specific investigation (necessary to find the correct questionnaire set belonging to this assessment). The questionnaire set contains the set of questionnaires which should be retrieved for the client, their order of retrieval and their URL or web address. The CA checks which of the questionnaires in this assessment is the first one that is not yet initiated or not fully completed and gets the web address of this questionnaire, upon which the client is redirected to this questionnaire.

The CA needs to send information about the corresponding assessment to the redirected questionnaire. Sending the session id is a rather obvious choice, but security wise this is not a responsible decision. Instead, the CA sends a GUID (Globally unique identifier) included in the questionnaire URL. The questionnaire can retrieve the correct assessment based on this GUID.

Once a questionnaire is completed or interrupted by the client, the questionnaire sends a redirect to the CA. In case of interruption by the client, the assessment will be terminated. In case the questionnaire was completed, the CA continues the above process of finding the next uninitiated or unfinished questionnaire with the lowest order. If there are no more questionnaires left for the assessment, the client is redirected to the complete page.

The CA was developed as an ASP.Net Web Application, using C# as the programming language.

2.5.4.1 Implementation

The layout of the application was HTML coded. Separate class files were used for C# code, regarding the dynamic contents of the pages. This approach allows a clean distinction between the presentation (HTML) code and the execution logic.

2.5.5 PCH professional Application (PA)

After a client has completed an assessment, the results will be discussed during a standard health examination. Therefore, the PCH professional must be able to log into the PCH professional Application (PA) to view the results.

The result of an assessment consists of a total problem score (TPS) and three sub-scale scores, graphically presented. Another presentation of the result consists in a listing of the administered items. Apart from the questions asked and answers given, the list also includes a measure of the seriousness of each response and the subscales to which each questions contributed.

The PA has to gather data from a data source and display it to the user. This process is more complex than it seems. First of all, only completed assessments should be displayed, for incomplete assessments will not provide reliable results. Furthermore, the questionnaire data is stored in a MR Database. This data is structured according to MR Interview questionnaires, not suitable to use without special interfaces. An importer is used to retrieve data from the MR Database, convert it to useful date for reporting, and storing it in the Application Database, ready for use by the PA.

2.5.5.1 Implementation

The layout of the PCH professional Application was HTML coded. Separate class files were used for C# code, regarding the dynamic contents of the pages. This approach allows a clean separation of the HTML code from the application logic.

The Importer is implemented as a standalone ASP.NET Web Application, written in C#. The application consists of several screens, listing projects, assessments, and results. The assessments' listing (see figure 12) contains a table with assessments, belonging to a specific investigation. The list can be sorted by client number, surname, and birth date. Furthermore, the listing can be filtered on (parts of) client number, surname, first name, and birth date. If the result of an assessment is available, the column "Rapportage beschikbaar" will contain the link "Ja", which needs to be clicked on to view the result.

The result consists of two separate views (see figures 12 and 13). The Score view illustrates the score with a graphical representation, containing four bars indicating the score. These bars were implemented using a custom graphics component in C#. The Items view shows a detailed listing of all the questions of that specific assessment. The items can be sorted by item number, response, one of the subscales, or impact (i.e. the seriousness of the answer given).

2.5.5.2 Results

The assessment is retrieved from the MR Database by the importer, converted, and then stored in the Application Database. To make sure the assessment data is available shortly after completing an assessment, the importer is triggered every 2 minutes to retrieve any changed data. In other words, the waiting time for assessment results to be available for reporting is at most 2 minutes after completion of the questionnaire.

2.5.6 Application Database

All application data is stored in the Application Database (AD). The main functions of the application have been discussed throughout this report, namely administering questionnaires and reporting assessment results. However, the system is also used for other, supporting or non-functional, purposes. For all these various purposes, the Application Database (AD) is an essential component. To understand what role that AD plays in different situations, consider the following overview of the system components and their dependency on the AD.

Client Application
- Client authorization
- Session information
- Items information

Questionnaire Component
- Session verification
- Respondent registration

PCH professional Application
- PCH professional authorization
- Project listing
- Assessment listing
- Child personal data

- Layout problem scale bar
- Session verification

Each component uses the AD for various reasons. Keeping that in mind, an initial data model was designed. The model started with a mapping of the objects to tables, such as "Client", "Item", and "Session". Then table attributes were determined, either extracted from the requirements program, or added after consultation with database experts. The next step was normalizing the database, preventing n-n relationships between tables.

2.5.6.1 Implementation

The database was modelled in Microsoft Visio. The database itself was created in SQL Server 2000, in compliance with the standard used by TNO.

Initially, a test database based on the data model was used during the implementation and testing of the listed system components. The tables were manually created. Some data was inserted through an automated process, such as DTS for copying the item bank from a text file to the database. Other tables were manually filled with data, such as the case of creating clients or questionnaire sets.

Once all the modules were tested and functioning, a database generating SQL script was created. This script was then executed to create a fresh, production version database, including all tables, their relationships, and attributes. The database also contains custom views and stored procedures.

2.5.7 Summary of the results of the application development

The goal of the application development was to build a system according to the requirements program. In other words, build a web based application for clients, managing online questionnaires, and also a web based application for PCH professionals, used for presenting the results of assessments. Also required was a data storage function. These were all realized, using various techniques and software. Although these subsystems resulted in independent modules, they were interconnected to form the IRT application.

The Client Application (CA) verifies authenticity, retrieves the related assessment, and seamlessly redirects the user to the corresponding questionnaires, as if an assessment is one large questionnaire. MR Interview questionnaires are available online, although only through the CA, and the IRT questionnaire results in expected assessments. The PCH professionals' Application verifies authenticity, and authorisation, listing assessments of associated clients, and the results belonging to each assessment.

3 Summary and Discussion

Aim of this study was to assess was whether IRT models could possibly offer a perspective on a fast, short, yet high quality identification of children with emotional and behavioural problems in community based Child Health Care services. We used a data set containing information about 2041 children, which we randomly split into two subsamples, a calibration sample, to assess the fit of the items and to determine the item locations, and the validation sample with which a Computerized adaptive test was simulated in order to assess the efficiency and quality of such a test.

First we assessed to what extent the items of the PSC, SDQ, PSYBOBA and CBCL fitted the Partial Credit Model. Only eleven items out of the 201 items involved did not fit the model and were removed from the analysis. The item locations the remaining 190 items were then determined. We calculated the persons' location (theta) on the latent scale, and determined a suitable cut-off point that would enable a correct classification, corresponding to a clinical CBCL score, yes or no.

In order to evaluate the efficiency and quality of an IRT based CAT the validation sample was used for a simulation study, in which the answers on paper and pencil questionnaires were used as if they were the responses on items in a CAT. We assessed to which extent the CAT procedure would enable a correct distinction, using two criteria: having a CBCL TPS score in the clinical range or less, and currently being treated for psychosocial problems or not. Due to the small number of cases in the validation sample, sensitivity indices could only be estimated very roughly. The results showed that the CAT procedure had a sensitivity of 89% for a CBCL TPS in the clinical range, with a specificity of 90%. Most children with borderline CBCL scores got an estimated person's location above the cut off point of -1.9. Prediction of current treatment status was worse, with a sensitivity of 55% and a specificity of 89%.

These results suggest that an IRT based CAT selects the right items to achieve a fast classification of children as probable cases, with a predefined confidence of 95%. For more than half of the children a maximum of ten items seems enough.

As the results of the analyses were very promising a working prototype, ready for a first field test, was developed. This prototype was developed as an internet application, consists of three main modules and offers the following functions:

Client Application
- Client authorization
- Administration of IRT-CAT driven questionnaires
- Administration of linear questionnaires

PCH professional Application
- PCH professional authorization
- Listing of investigation associated with the PCH professional
- Lisiting of available assessments
- Graphical presentation of scale scores, distinguishing between scores in the normal, borderline and clinical range
- Presentation of questions asked and answers given

3.1.1 Discussion

Limitations

The results of the present evaluation study have to be interpreted carefully. First of all, the validation sample is quite small and the number of cases in this sample is very small. The 95% confidence intervals of the calculated sensitivity indices is therefore large. Furthermore, the 'true' classification, based on the CBCL score, is defined using the same items that the simulation uses. Therefore one might conclude that the calculated sensitivity and specificity indices are an overestimation. However, in this study our aim was not to validate a CAT against some true score for emotional and behavioural problems, but to assess to which extent a CAT could replicate the distinction made by the CBCL between a clinical or lower score, with fewer items than the original questionnaires. The parameters we present simply indicate how well the CAT, using a limited number of items, replicates the CBCL classification.

We chose a cut off point based on the calibration sample and tested this in the validation sample. The results showed that a CAT, using this cut off point, could replicate the CBCL classification with few questions and with a high sensitivity and specificity. Yet, it is clear that the performance of a CAT and the chosen cut off pint depend of the prevalence of problems in the sample investigated. Both the calibration sample and the validation sample are a random selection from a sample which may be considered as representative for the population – in this age group – under care off PCH. In other groups the cut off points needs reconsideration.

The data set we used for the CAT simulations is incomplete, because respondents only took either the PSC or the SDQ or the PSYBOBA together with the CBCL. The CAT simulations, of course, could only select from items that were actually administered. When the CAT is to be applied in real life, all items will be available for selection and this will probably lead to faster convergence and more accurate results.

Finally, the evaluation in this phase could only be done by means of simulations. The results look very promising. However, it is not certain if parents answering questions in a CAT will respond in the same way as they do on questions in a paper and pencil questionnaire. In paper and pencil questionnaires, for example, they will get more questions, sometimes on similar or similarly looking questions. They may then tend to deny the existence of specific problems, thinking they already reported on them when answering an earlier, more or less similar question. Ultimately, therefore, the quality of a CAT on psychosocial problems will have to be established in a study in which parents are actually asked to answer a real life CAT.

Fit with the literature

IRT based CAT procedures have been widely used in the field of intelligence testing and the assessment of school achievements (ref) and these studies are promising in that they show that using such methods can lead to accurate testing with far fewer items than in tests based on more traditional psychometrics (refs). More recently studies on Health Related Quality of Life have shown that in this field too, IRT based techniques are promising, despite the fact that one of the assumptions of most IRT models (unidimensionality) is a more problematic one here. Studies on the application or IRT models in the field of the identification of psychosocial problems are just beginning to appear. Gardner{Gardner, 2002 550 /id} evaluated the extent to which a

multidimensional adaptive test could be used to replicate screening decisions based on the Pediatric Symptom Checklist, a question quite similar to ours.

In order to assess the correspondence between the Adaptive PSC and the original 35 items PSC he calculated kappa. The kappa he found (0.84) was higher than the corresponding figure we found. The mean number of items he needed to achieve this was 12 items, out of 35. It is not exactly clear why he found a much higher correspondence than we did. One explanation might be that he limited himself to PSC items only, while we used items of four questionnaires in trying to replicate a case definition based on one of them. We therefore have less overlap between items in the CAT and items that are used for the criterion. Perhaps, also, our cut-off point should be somewhat lower (to achieve an optimal kappa) than the cut off point we decided to use in this study, as most children with a borderline CBCL TPS score got an estimated person's location above this cut off point.

Another explanation may be that Gardner used a multidimensional model. We decided to evaluate the quality of a single dimensional model, because of the fact that the CBCL in calculating a single total problem score and then using cut-off point based on this single score, assumes unidimensionality too, despite the existence of CBCL subscales. The question whether incorporating subscales in a CAT would add to the quality and / or efficiency will be addressed in a following study.

Conclusion

Despite this study's limitation, it is clear that IRT based CAT is an option worthwhile to explore. The results of this study were used to develop a system, with which IRT based CAT can be evaluated in the actual context of Preventive Child Health Care. A working prototype of such a system allows for such a field test to be carried out.

4 Reference List

Achenbach, T. M. & Rescorla, L. A. (2001). Manual for the ASEBA School-Age Forms & Profiles. Burlington VT: University of Vermont, Research Center for Children, Youth and Families.

Anderson, D. L., Spratt, E. G., Macias, M. M., Jellinek, M. S., Murphy, J. M., Pagano, M. et al. (1999). Use of the pediatric symptom checklist in the pediatric neurology population. Pediatr Neurol., 20, 116-120.

Andric, D. and Luo, G. (2003). Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. J Appl Meas, 4, 205-221.

Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. J.Am.Acad.Child Adolesc Psychiatry., 44, 557-564.

Brugman, E., Reijneveld, S. A., Verhulst, F. C., & Verloove-Vanhorick, S. P. (2001). Identification and management of psychosocial problems by preventive child health care. Arch Pediatr Adolesc Med, 155, 462-469.

Costello, E. J. & Janizewski, S. (1990). Who gets treated? Factors associated with referral in children with psychiatric disorders. Acta Psychiatr.Scand., 81, 523-529.

Durlak, J. A. & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: a meta-analytic review. Am.J.Community Psychol., 25, 115-152.

Eden van, J., Rover, C. d., Verhappen, M. J., & Vries-Lequin, I. d. (2003). PSYBOBA. A pilot study on the usability of a questionnaire for psychosocial problems among primary school children aged 7 till 12 [in Dutch]. GGD Nederland.

Fliege, H., Becker, J., Walter, O. B., Bjorner, J. B., Klapp, B. F., & Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT). Qual.Life Res., 14, 2277-2291.

Gardner, W., Kelleher, K. J., & Pajer, K. A. (2002). Multidimensional adaptive testing for mental health problems in primary care. Med Care, 40, 812-823.

Gardner, W., Shear, K., Kelleher, K. J., Pajer, K. A., Mammen, O., Buysse, D. et al. (2004). Computerized adaptive measurement of depression: a simulation study. BMC.Psychiatry, 4:13., 13.

Glascoe, F. P. (1999). Using parents' concerns to detect and address developmental and behavioral problems. J.Soc.Pediatr Nurs., 4, 24-35.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. Br.J.Psychiatry, 177, 534-539.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2003). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. Int.Rev.Psychiatry, 15, 166-172.

Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. Eur.Child Adolesc Psychiatry., 9, 129-134.

Goodman, R. & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? J.Abnorm.Child Psychol., 27, 17-24.

Guillemin, F. (1995). Cross-cultural adaptation and validation of health status measures. Scand.J.Rheumatol., 24, 61-63.

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J.Clin.Epidemiol., 46, 1417-1432.

Horowitz, L. M., Wang, P. S., Koocher, G. P., Burr, B. H., Smith, M. F., Klavon, S. et al. (2001). Detecting suicide risk in a pediatric emergency department: development of a brief screening tool. Pediatrics, 107, 1133-1137.

Jellinek, M., Evans, N., & Knight, R. B. (1979). Use of a behavior checklist on a pediatric inpatient unit. J.Pediatr, 94, 156-158.

Jellinek, M., Little, M., Murphy, J. M., & Pagano, M. (1995). The Pediatric Symptom Checklist. Support for a role in a managed care environment. Arch Pediatr Adolesc Med, 149, 740-746.

Jellinek, M., Murphy J.M., Robinson J., Feins A., Lamb, S., Fenton T. et al. (1988). Pediatric Symptom Checklist: screening schoolage children for psychosocial dysfunction. J.Pediatr, 112, 201-209.

Jellinek, M. S. & Murphy, J. M. (1988). Screening for psychosocial disorders in pediatric practice. Am.J.Dis.Child, 142, 1153-1157.

Jutte, D. P., Burgos, A., Mendoza, F., Ford, C. B., & Huffman, L. C. (2003). Use of the Pediatric Symptom Checklist in a low-income, Mexican American population. Arch Pediatr Adolesc Med, 157, 1169-1176.

Kelleher, K. J., McInerny, T. K., Gardner, W. P., Childs, G. E., & Wasserman, R. C. (2000). Increasing identification of psychosocial problems: 1979-1996. Pediatrics, 105, 1313-1321.

Lavigne, J. V., Binns, H. J., Christoffel, K. K., Rosenbaum, D., Arend, R., Smith, K. et al. (1993). Behavioral and emotional problems among preschool children in pediatric primary care: prevalence and pediatricians' recognition. Pediatric Practice Research Group. Pediatrics, 91, 649-655.

Linacre, J. M. (2002). Rasch Measurements Transactions, 16, 878.

Moffitt T.E. & Caspi A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. Development and Psychopathology, 355-375.

Murphy, J. M., Arnett, H. L., Bishop, S. J., Jellinek, M. S., & Reede, J. Y. (1992). Screening for psychosocial dysfunction in pediatric practice. A naturalistic study of the Pediatric Symptom Checklist. Clin.Pediatr (Phila), 31, 660-667.

Reijneveld, S. A., Vogels, A. G., Brugman, E., van Ede, J., Verhulst, F. C., & Verloove-Vanhorick, S. P. (2003). Early detection of psychosocial problems in adolescents: how useful is the Dutch short indicative questionnaire (KIVPA)? Eur.J.Public Health, 13, 152-159.

Revicki, D. A. & Cella, D. F. (1997). Health status assessment for the twenty-first century: item response theory, item banking and Computerized adaptive testing. Qual.Life Res., 6, 595-600.

Verhulst, F. C., Dekker, M. C., & Van der, E. J. (1997). Parent, teacher and self-reports as predictors of signs of disturbance in adolescents: whose information carries the most weight? Acta Psychiatr.Scand., 96, 75-81.

Verhulst, F. C. & Ende van der, J. (1997). Factors associated with child mental health service use in the community. J.Am.Acad.Child Adolesc Psychiatry, 36, 901-909.

Verhulst, F. C. & Koot, H. M. (1996). Manual for the CBCL/4-18 [in Dutch]. Rotterdam: EUR/Sophia Kinderziekenhuis.

Verhulst, F. C. & Van der, E. J. (1992). Six-year stability of parent-reported problem behavior in an epidemiological sample. J.Abnorm.Child Psychol., 20, 595-610.

Verhulst, F. C., Van der, E. J., & Rietbergen, A. (1997). Ten-year time trends of psychopathology in Dutch children and adolescents: no evidence for strong trends. Acta Psychiatr.Scand., 96, 7-13.

Vestering, N. & Buuren, S. v. (2004). Two New Bayesian item selection methods and their application inComputerized Classification Testing. Leiden: TNO Quality of Life.

Vogels, A. G. C., Crone, M. R., Hoekstra F., & Reijneveld, S. A. (2005). Drie vragenlijsten voor het opsporen van psychosociale problemen bij kinderen van zeven tot twaalf jaar Leiden: TNO Kwaliteit van leven.

Vogels, T., Reijneveld, S. A., Brugman, E., Hollander-Gijsman, M., Verhulst, F. C., & Verloove-Vanhorick, S. P. (2003). Detecting psychosocial problems among 5-6-year-old children in preventive Child Health Care: the validity of a short questionnaire used in an assessment procedure for detecting psychosocial problems among children. Eur.J.Public Health, 13, 353-360.

Wainer, H. (1990). Computerized Adaptive Testing. A Primer. Hillsdale: Lawrence Erlbaum Associates, Publishers.

Wiefferink, C. H., Reijneveld, S. A., de Wijs, J., Swagerman, M., Campman, D., & Paulussen, T. G. (2006). Screening for psychosocial problems in 5-6-year olds: a randomised controlled trial of routine health assessments. Patient.Educ.Couns., 60, 57-65.

Wright, D. & Masters, G. (1982). Rating Scale Analysis. Institute for Objective Measurement.

A Programma van Eisen

Introductie

Dit document beschrijft het programma van eisen (PvE) van het project CAT Psychosociale problematiek jeugd 7-12 jaar. In het PvE worden de *requirements* (eisen) gedefinieerd en gespecificeerd. Dit proces wordt ook wel *Requirements Engineering* genoemd. Requirements engineering bestaat uit twee hoofd activiteiten, namelijk **requirements elicitation** en **analyse**. Beide activiteiten zijn gefocust op het vormen van een beeld van het systeem, gezien vanuit het oogpunt van de gebruikers. Hoe deze taken zijn uitgevoerd en wat de uitkomsten zijn worden uiteindelijk gespecificeerd in dit requirements document.

Requirements elicitation (verwerven) is gebaseerd op het beschrijven van het doel van het systeem. De opdrachtgever (klant), de ontwikkelaars en de gebruikers identificeren het probleemgebied en definiëren zodoende een systeem dat van toepassing is op het probleem. Requirements elicitation definieert een **systeemspecificatie** die gezien kan worden als een contract tussen de opdrachtgever (klant) en de ontwikkelaars. Deze systeemspecificatie wordt tijdens de analyse gestructureerd en geformaliseerd tot een **analysemodel**.

De systeemspecificatie en het analysemodel presenteren dezelfde informatie over het systeem. Het verschil tussen beide zit onder andere in de gebruikte taal en notatie. De systeemspecificatie wordt geschreven in een natuurlijke taal en het analysemodel wordt gepresenteerd in een meer formele schrijfwijze. De systeemspecificatie ondersteunt de communicatie tussen alle bij het systeem betrokken partijen, terwijl het analysemodel de onderlinge communicatie tussen de ontwikkelaars bevordert.

Er zijn verschillende standaarden voor het opstellen van een requirements document. In dit geval is gekozen om gebruik te maken van de IEEE / ANSI 830-1998 standaard voor requirements specificatie. Uit deze standaard worden de stabiele (vaste) onderdelen overgenomen, terwijl de variabele onderdelen worden aangepast aan het systeem dat voor deze opdracht wordt ontwikkeld. Het volgens deze standaard ontwikkelde document wordt ook wel *Software Requirements Specification* (SRS) genoemd.

In dit hoofdstuk wordt gekeken naar het doel van dit requirements document en de afbakening van het systeem. Verder worden vaak gebruikte termen uitgelegd, zodat de tekst ook leesbaar is voor mensen zonder enige kennis van de opdracht of het behandelde onderwerp. Tenslotte wordt een overzicht van de resterende hoofdstukken gegeven.

Doel van het PvE

Het Programma van Eisen (PvE) is een formeel document, waarin de eisen van een op te leveren systeem worden gedocumenteerd. Hierin worden de criteria gespecificeerd waaraan het uiteindelijke systeem zal worden getoetst. Het PvE geeft een overzicht van de eisen aan alle stakeholders van het betreffende project, zoals de opdrachtgevers, systeemontwikkelaars, projectmanagers en gebruikers. Het kan gezien worden als een contract, dat precies vastlegt wat de opdrachtgever en de ontwerpers zijn overeengekomen.

Het requirements document of SRS wordt voor verschillende doeleinden gebruikt door de betrokken partijen, namelijk:

- Systeemontwikkelaars: Als input voor het te ontwikkelen systeem. Het document verschaft een indicatie over de vereiste inzet.
- Projectmanagers: Om te controleren of de requirements voldoen aan de door hun gestelde eisen. Het wordt ook gebruikt voor projectplanning.
- Gebruikers: Om feedback te geven over de requirements die kunnen leiden tot eventuele wijzigingen.
- Opdrachtgever: Om een overzicht te krijgen van alle requirements (ontbreken er geen requirements?).

Afbakening

De te ontwikkelen applicatie krijgt de voorlopige naam van TNO CAT Vragenlijst Engine.

Deze op IRT gebaseerde CAT applicatie moet zorgen voor een snelle en betrouwbare afname van vragenlijsten voor het opsporen van kinderen met psychosociale problematiek. Het systeem bepaalt voor (ouders van) kinderen – de cliënten – aan de hand van bepaalde parameters welke vragen gesteld zullen worden. Na de afname kan een jeugdarts of jeugdverpleegkundige (JAV) aan de hand van de rapportage een diagnose vaststellen.

De requirements zoals opgesteld in dit document zijn alleen geldig voor een werkend prototype dat zal worden gebruikt voor de testfase binnen het IRT-CAT vragenlijst engine project. Voor de uiteindelijke versie van het systeem zijn mogelijk additionele requirements vereist, die voortvloeien uit de eventuele wensen van de opdrachtgever. Een deel van die additionele requirements wordt onder de kop *Additionele Requirements* geplaatst. Deze additionele requirements moet men echter niet als definitief beschouwen, aangezien ze nog op haalbaarheid getoetst moeten worden.

Definities, acroniemen en afkortingen

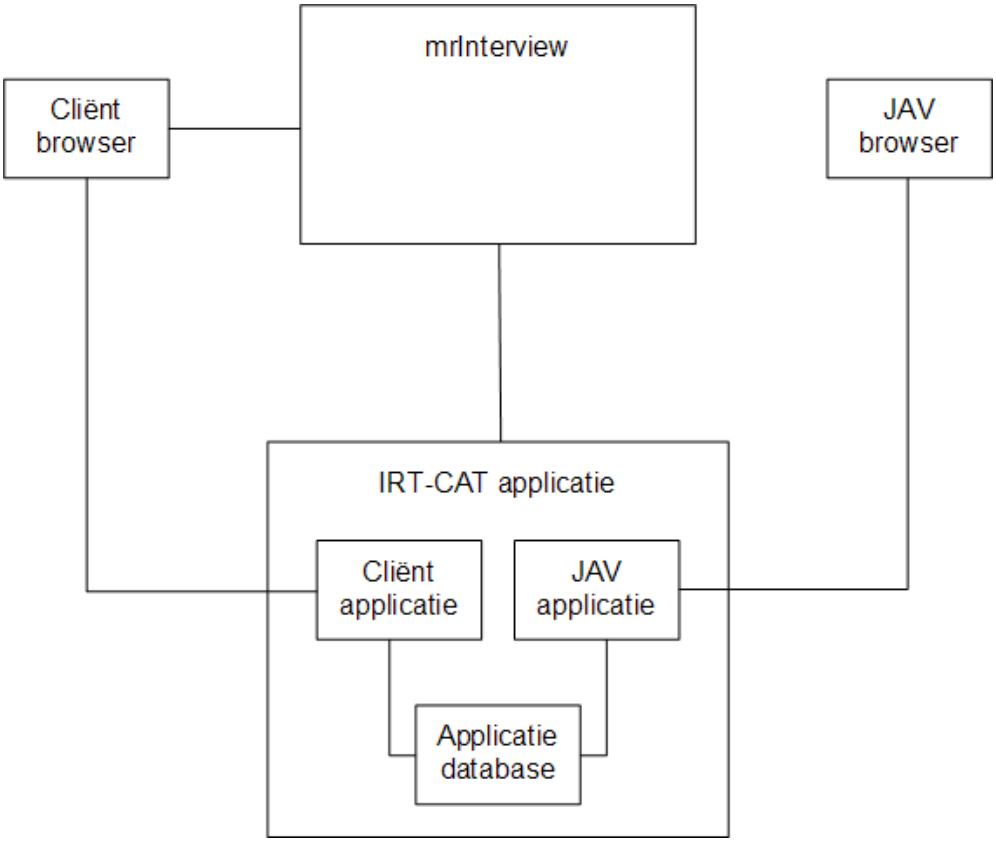
Beheerder	Er wordt onderscheid gemaakt tussen centrale en lokale beheerder. De centrale beheerder onderhoudt o.a. de itembank. Hij geeft onder andere aan welke items bij een afname gesteld mogen worden. De lokale beheerder zorgt o.a. voor de user accounts. In de pilotfase zal TNO fungeren als zowel centrale als lokale beheerder. In de post-pilotfase wordt de rol van lokale beheerder overgenomen door JGZ.
Cliënt	Dit zijn ouders van kinderen die de vragenlijst invullen. Zij hebben geen toegang tot cliënt- of afnamegegevens.
JAV	De jeugdarts en/of jeugdverpleegkundige (JAV) belast met het beoordelen van de afgenomen vragenlijst. Hij heeft toegang tot afname- en cliëntgegevens.
CAT	Computerized Adaptive Testing (CAT) biedt de mogelijkheid om een selectie te maken van vragen die gesteld zullen worden.
IRT	Item Response Theory (IRT) bepaalt de kans op een positief antwoord voor items door gebruik te maken van item parameters en eigenschappen van de respondenten.
GGD	Gemeentelijke Gezondheidsdienst (GGD)
JGZ	Jeugdgezondheidszorg (JGZ)
UMTS	Universal Mobile Telecommunications System (UMTS) is de verbeterde, snelle versie van het huidige GSM/GPRS-netwerk. Gebruikers kunnen via een laptop met een UMTS internetverbinding gebruik maken van het afnamesysteem (voor de afname en de rapportage).
Wbp	De Wet bescherming persoonsgegevens (Wbp) biedt bescherming bij de registratie van persoonsgegevens, alsook bij het totale proces van verwerken van die gegevens.

Leeswijzer

In hoofdstuk 2 wordt de algemene omschrijving van het afnamesysteem behandeld. Vervolgens worden in hoofdstuk 3, Specifieke Requirements, de requirements specificatie en het analyse model van het nieuwe systeem (in de pilotfase) gedocumenteerd. Tenslotte worden de additionele requirements van het op te leveren systeem (in de post-pilotfase) in hoofdstuk 4 gespecificeerd. Deze requirements zullen in overleg met de opdrachtgever na de pilotfase geïmplementeerd worden.

Algemene beschrijving

In dit hoofdstuk worden de factoren die invloed hebben op (de eisen van) het systeem behandeld. Hier worden geen specifieke requirements beschreven. Er wordt slechts achtergrond informatie gegeven over de requirements die in het volgende hoofdstuk worden behandeld.



Figuur 1 - Klassendiagram

Beschrijving systeemobjecten:

MrInterview
Dit is de VragenafnameApplicatie. Deze applicatie bepaalt aan de hand van eerder gestelde vragen welke vragen nog gesteld kunnen worden.

Cliënt browser
User interface van de cliënt (respondent)

JAV Browser
User interface van de JAV

Applicatiedatabase
Hier worden onder andere de cliënt- en JAV gegevens opgeslagen. Verder bevat de database ook gegevens over items (zoals de threshold parameters).

IRT-CAT Applicatie (Cliënt)
De cliënt meldt zich aan bij de cliëntapplicatie. Hier vindt authenticatie plaats. De cliëntgegevens worden vergeleken met data in de applicatiedatabase. Slechts als de authenticatie slaagt, wordt de cliënt doorverwezen naar de mrInterview vragensessie.

IRT-CAT Applicatie (JAV)
De cliënt meldt zich aan bij de JAV applicatie. Hier vind authenticatie plaats. De JAV gegevens worden vergeleken met data in de applicatiedatabase. Na een geslaagde authenticatie mag de JAV gegevens van MrInterview opvragen (zoals rapportages).

Specifieke requirements

In dit requirements document worden de software en systeem requirements gedocumenteerd. Dit hoofdstuk behandelt de functionele en niet-functionele requirements die bij het systeem behoren. Deze requirements zijn bestemd voor het op te leveren systeem in de pilotfase, op de met grijs gemarkeerde eisen na. Die zullen pas na de pilot in het systeem geïmplementeerd worden..

Functionele requirements

De functionele requirements beschrijven de implementatie onafhankelijke interacties tussen het systeem en de directe omgeving. Tot de omgeving worden gerekend de gebruiker en alle andere externe systemen waarmee het systeem communiceert.

De functionele requirements hebben betrekken op de hoofdprocessen van het systeem, namelijk: afname en rapportage.

I AFNAME
I.1 Het systeem zal per vragenlijst een korte en duidelijke instructie geven aan de respondent.
I.2 Het systeem maakt beantwoording met muis en (eventueel) met toetsenbord mogelijk (letters of nummers).
I.3 Het systeem zal gebruik maken van een logische volgorde van tabstops en de spatiebalk bij het beantwoorden.
I.4 Het systeem zal in verband met mogelijke sekse- of leeftijdsgerelateerde verschillen in normen van vragenlijsten de geboortedatum en sekse vragen (geen IRT vraag). Deze gegevens worden tijdens de rapportage getoond.
I.5 Het systeem maakt gebruik van een IRT-CAT algoritme, waarmee aan de hand van de response op een item het volgende te selecteren item wordt bepaald.
I.6 Met behulp van het IRT-CAT algoritme kan de vaardigheid van personen worden bijgehouden. Hiermee wordt de *ability estimation* berekend.
I.7 Het IRT-CAT algoritme haalt de threshold waarden (berekend met behulp van een IRT programma) van alle benodigde items uit de Applicatie database.
I.8 Naast de IRT gestuurde items moeten ook items zonder IRT-sturing afgenomen kunnen worden. De resultaten zullen worden gebruikt voor de evaluatie van de ClientApplicatie, door vragen over cliëntervaringen te stellen.
I.9 Het systeem maakt de afname mogelijk van meerdere vragenlijsten.
I.10 Het systeem maakt het mogelijk om een afname te definiëren als een combinatie van vragenlijsten.
I.11 Door op de knop ‘verder’ te drukken wordt de volgende vraag gesteld (indien meer vragen nodig zijn).
I.12 Het systeem zal de afname beëindigen wanneer:
I.12.1 deze geresulteerd heeft in een voldoende nauwkeurige meting van de scores op de Totale Probleemscore, Externaliseren, Internaliseren en Hyperactief gedrag;
I.12.2 wanneer het maximum aantal vragen is voorgelegd; maximaal 15 vragen uit overall schaal en maximaal 8 vragen per subschaal;

I.12.3 er gedurende een periode van 10 minuten geen vragen worden beantwoord (time out).

I.13 Ouders kunnen de sessie vroegtijdig beëindigen, waarbij het systeem vraagt of de gegeven antwoorden opgeslagen en door de JGZ ingezien mogen worden; bij een negatief antwoord wordt de koppeling naar persoonsgegevens verwijderd.

II RAPPORTAGE

II.1 Uitslagen worden alleen aan de jeugdarts/-verpleegkundige (JAV) gerapporteerd.

II.2 De voor de JAV beschikbare afnamen worden in een lijst gerapporteerd. Per afname worden getoond: cliëntnummer[2], voor- en achternaam[3], geslacht, geboortedatum en afnamedatum.

II.3 Er kan gesorteerd worden op cliëntnummer, geboortedatum en afnamedatum.

II.4 Er kan gezocht worden naar specifieke afnamen. Daarvoor kunnen de zoekvelden cliëntnummer en geboortedatum worden gebruikt. Als tenminste één van de zoekvelden is ingevuld, verschijnt na het drukken op de zoekknop een lijst met afnamen die aan de criteria voldoet.

II.5 Door klikken op een afname (meting) wordt de rapportage van een specifiek kind getoond.

II.6 Bij de rapportage van individuele gegevens wordt op elk scherm en -afdruk getoond: voor- en achternaam, cliëntnummer, geboortedatum en geslacht van het kind, inclusief de afnamedatum.

II.7 Het systeem toont de volgende meetgegevens in het rapportagescherm: de totale probleem score (vergelijkbaar met die van de CBCL en de SDQ) met behulp van schaalscores en subschaalscores op externaliserend, internaliserend en hyperactief gedrag [en mogelijk op faalangst en depressiviteit], in relatie tot afkappunten borderline en klinisch indien die voor de betreffende vragenlijst beschikbaar zijn.

II.8 Het systeem toont de scores aan de hand van horizontale staafdiagrammen, met de kleuren groen, oranje en rood voor achtereenvolgens de gebieden normaal, borderline en klinisch.

II.9 Het systeem beschikt in het rapportagescherm over een knop voor een uitgebreide rapportage, waardoor de gestelde vragen, de gegeven antwoorden en het probleemgehalte van elk gegeven antwoord worden getoond.

II.10 Indien meerdere metingen per kind beschikbaar zijn, worden die getoond op volgorde van afname, waarbij gekozen kan worden voor de rapportage van een van de metingen.

II.11 Er wordt vermeld hoe de afname werd afgesloten, dus door het systeem (bereiken van time out, maximaal aantal vragen of voldoende nauwkeurigheid) of door de respondent.

II.12 Rapportages moeten geprint kunnen worden.

Niet-functionele requirements

Niet-functionele requirements kunnen gedefinieerd worden als te zijn de beperkingen van het systeem. Deze requirements beschrijven de zichtbare aspecten van het systeem welke niet direct gerelateerd zijn aan het functioneren van het systeem.

---

[2] Het cliëntnummer is een alfanumerieke code, zodat de mogelijkheid bestaat voor het gebruik van een eigen cliëntcodering door een GGD, of het BSN (indien beschikbaar).

[3] Tijdens de testfase zijn de voor- en achternaam niet beschikbaar, omdat deze velden niet worden ingevuld. Na de testfase mogen de artsen bij het aanmaken van nieuwe cliënten deze wel invullen.

III OPERATIONELE REQUIREMENTS

III.1 Het programma moet kunnen draaien in een internetomgeving.

III.2 Afname van de vragenlijst is thuis bij de cliënt mogelijk via eigen internetverbinding.

III.3 Cliënten kunnen op locatie bij een GGD gebruik maken van een laptop met een UMTS (indien nodig) internetverbinding.

III.4 De rapportage aan JAV dient op de JAV-pc beschikbaar te zijn op standaardcontactmomenten tussen JGZ en (ouders van) kinderen, zoals tijden het Periodiek Gezondheidsonderzoek (PGO). Hiervoor beschikt de JAV over een laptop met internetverbinding via UMTS (indien nodig).

IV OPSLAG REQUIREMENTS

IV.1 Export van opgeslagen gegevens door de beheerder naar een algemeen formaat bv. csv of SPSS .sav onder andere ten behoeve van opname in ECD of analyses op groepsniveau. De keuzemogelijkheden zijn selectie van leeftijd, sekse, soorten vragenlijsten, postcode en eventueel school/groepsniveau. Verder kunnen ook JGZ gerelateerde gegevens, zoals dienst en arts worden opgeslagen.

IV.2 De JAV-gegevens en cliëntgegevens worden in de Applicatie database opgeslagen. Deze gegevens moeten aan elkaar gekoppeld zijn. De koppelingsdata zullen worden opgeslagen in aparte tabellen.

IV.3 De automatiseringsafdeling van KvL maakt back-ups van data.

IV.4 De cliëntdata wordt na de pilot vernietigd.

V PERFORMANCE REQUIREMENTS

V.1 Het programma sluit aan bij standaard Microsoft Windows conventies

V.2 De minimale systeemeisen voor ouders die de gegevens thuis invullen zijn: Windows 98 met Internet Explorer versie 5 of hoger, schermresolutie 800x600 en een voldoende snelle verbinding (tenminste 56 k).

V.3 De wachttijd na het selecteren van een antwoord en voor het verschijnen van de volgende vraag duurt maximaal één seconde.

VI QUALITY REQUIREMENTS

VI.1 Usability wordt getest in de pilot (o.a. door TNO TM).

VI.2 Software wordt ontwikkeld volgens NEN 7510 standaard voor informatiebeveiliging.

VII SECURITY REQUIREMENTS

Om ongeautoriseerde toegang tot het systeem te voorkomen en de integriteit van de data te waarborgen bij eventuele schade, worden de volgende eisen gesteld:

VII.1 Voor de afname is een gebruikersnaam en een wachtwoord vereist.

VII.2 Voor de rapportage is een gebruikersnaam en een wachtwoord vereist.

VII.3 Gebruikersnamen en wachtwoorden worden per GGD verstrekt bij uitnodiging voor een PGO.

VII.4 Gebruikersnamen en wachtwoorden worden gegenereerd door de centrale beheerder (in de pilot door TNO).

VII.5 Identificatie bij GGD dient te geschieden door middel van de koppeling van een persoonscode aan een (elektronische of papieren) register.
VII.6 Alleen een JAV kan een rapportage van afnameresultaten opvragen (toegang door middel van autorisatie met de juiste gebruikersnaam/wachtwoord combinatie).
VII.7 Een JAV kan alleen rapportages van zijn eigen cliënten opvragen.
VII.8 In verband met afname via Internet worden standaard geen direct op individuele personen herleidbare gegevens gevraagd of opgenomen in de database.
VII.9 Een JAV en/of diens leidinggevende moeten op verzoek van direct betrokkene gegevens uit elk persoonsregister kunnen verwijderen, zoals vastgesteld door Wbp.

Pseudo Requirements

De pseudo requirements beschrijven door de opdrachtgever opgelegde ontwerp en implementatie beperkingen aan het systeem. Deze beperkingen kunnen zijn het gebruiken van een specifieke programmeertaal of database management system.
VIII.1 De cliënt- en JAV applicatie worden ontwikkeld in C# als ASP.NET webapplicaties, conform de door TNO Kwaliteit van Leven (KvL) gehanteerde standaard.
VIII.2 De vragenlijsten worden ontwikkeld in mrStudio (met mrInterview component).
VIII.3 De algoritmemodule wordt geschreven in C# en (mrScriptBasic).
VIII.4 De Applicatiedatabase is een SQL Server database.

Additionele Requirements (testfase)

IX.1 Toekomstige functies moeten slechts beschikbaar zijn voor daartoe bevoegde personen. Dat geldt bijvoorbeeld voor export naar ECD, export van gegevens ten behoeve van analyse op groepsniveau, beheer van persoonsregister, beheer van gegevensbestanden, beheer van vragenlijsten e.d.
IX.2 De toegangsrechten van de JAV worden per GGD zelf bepaald. Hiertoe kan een GGD voor iedere JAV aangeven welke cliëntgegevens hij mag inzien (zoals alleen zijn cliënten, iedereen binnen een bepaalde regio of provincie, etc.).
IX.3 Het programma moet ook op scholen gebruikt kunnen worden, ten behoeve van afname bij kinderen en leerkrachten en indien nodig samen met ouders van kinderen.
IX.4 De afnameresultaten van de uitgebreide rapportage kunnen gesorteerd worden op afnamevolgorde en probleemgehalte van items.
IX.5 Het systeem toont ook rapportages op school- of groepsniveau, met als uitslag de gemiddelde score van de klas en het percentage kinderen per gebied met bepaalde stoornissen (niet voorzien in eerste pilot).
IX.6 Het persoonsnummer moet gekoppeld kunnen worden aan het GGD-systeem. Hierbij is het van belang dat er een matching plaatsvindt tussen een persoonsnummer en de betreffende persoon.
IX.7 Verstrekken van gebruikersnaam en wachtwoord geschiedt op GGD-niveau (vergelijk eis VI.5).
IX.8 Het systeem toont een longitudinale rapportage indien er meerdere metingen per kind beschikbaar zijn.

Systeem modellen

De analyse levert een model van het systeem op: het analysemodel. Dit model is correct, compleet, consistent en verifieerbaar. Tijdens de analyse ligt de nadruk op het structureren en formaliseren van de requirements verkregen tijdens requirements elicitation. Het resulterende analysemodel zal worden gebruikt voor verificatie van de systeemspecificatie. Verder is dit model een nuttig middel om complexe problemen vroegtijdig in de ontwikkelfase te identificeren en op te lossen.

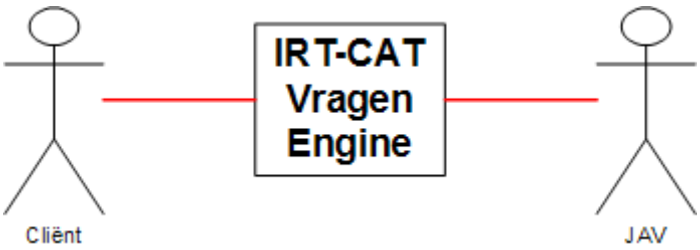
Het analysemodel bestaat uit drie afzonderlijke modellen:
- het functionele model (gepresenteerd in use case, scenario's)
- het analyse objectmodel (gepresenteerd in klasse diagrammen)
- het dynamische model (gepresenteerd in sequentie-, toestanddiagrammen)

Voor de modellering is gebruikt gemaakt van de diagram techniek UML (Unified Modelling Language [OMG, 1998]).

Functioneel model

In deze paragraaf zullen de requirements worden vastgesteld en beschreven aan de hand van actoren, scenario's en use cases. Hierdoor ontstaat de complete functionele specificatie van het systeem.

Actoren
Actoren zijn alle externe entiteiten die met het systeem communiceren. Een actor is een rol abstractie en niet noodzakelijk gekoppeld aan één persoon. Een persoon kan op verschillende momenten verscheidene rollen vervullen (cliënt of JAV). Aangezien beide rollen gebruik maken van verschillende functionaliteiten (en interfaces) van het systeem, worden deze twee rollen gemodelleerd als twee aparte actoren.



Figuur 2 – Actoren van de IRT-CAT Vragen Engine

Scenario's
Nadat de actoren van het systeem zijn vastgesteld, is het belangrijk om te bepalen welke functionaliteit toegankelijk is voor iedere actor. Dit wordt gedaan met behulp van scenario's.

Een scenario is een beschrijving van de wisselwerking tussen een actor en het systeem. Belangrijke eigenschappen van een scenario zijn concreetheid en duidelijkheid. Een scenario beschrijft een enkele functionaliteit van het systeem zoals ervaren door een actor. Het scenario wordt verder gespecificeerd in natuurlijke taal, zodat het goed leesbaar en bovendien gemakkelijk te begrijpen is, dus ook voor mensen met geringe kennis van IRT of CAT.

Scenario: vragenBeantwoord

Tabel 4.1 vragenBeantwoord scenario voor de StartVragenSessie use case.

Scenario naam	vragenBeantwoord
Actor instantie	Tim: Cliënt
Gebeurtenissen	1 Tim wil zich aanmelden voor een interview. Hij typt de url van de cliëntapplicatie in zijn browser. Hij komt vervolgens op het beginscherm. Daar vult hij de persoonscode en wachtwoord van het betreffende kind in en wacht vervolgens op instructies. 2 De cliëntapplicatie begint met de controle van de gegevens. Hij zoekt in de applicatiedatabase naar dezelfde persoonscode/wachtwoord combinatie. De combinatie bestaat en de applicatie verwijst Tim door naar de VragenafnameApplicatie. 3 Tim wordt doorgestuurd naar de startpagina van MrInterview. 4 MrInterview detecteert dat Tim toestemming vraagt om vragen te beantwoorden. MrInterview controleert de gegevens van Tim in de applicatiedatabase en haalt extra informatie op met betrekking tot de afname en autorisatie. Tim krijgt toestemming om te beginnen. 5 Voor Tim verschijnt de eerste vraag. Hij beantwoordt deze en gaat vervolgens door naar de volgende vraag. 6 Tim beantwoordt nog *n* (n<=15) vragen. 7 Het systeem heeft berekend dat er genoeg vragen zijn beantwoord om te kunnen beoordelen met de vereiste betrouwbaarheid. Het systeem laat Tim weten dat de vragenafname wordt gestopt. 8 Tim krijgt de mededeling dat de afname is geëindigd. Hij wordt bedankt voor zijn deelname en krijgt instructies over de verder te nemen stappen.

Scenario: afnameResultaten

Tabel 4.2 Het afnameResultaten scenario voor de ToonRapportage use case.

Scenario naam	afnameResultaten
Actor instantie	Mark: JAV
Gebeurtenissen	1 Mark wil een rapportage opvragen van een eerder afgenomen vragenlijst. Hij typt de url van de JAV-applicatie in zijn browser. Hij komt vervolgens op het beginscherm. Daar vult hij zijn gebruikersnaam en wachtwoord in. 2 De JAV-applicatie begint met de controle van de gegevens. Deze zoekt in de applicatiedatabase naar dezelfde gebruikersnaam / wachtwoord combinatie. De combinatie bestaat en de applicatie toont een scherm met afnameresultaten. Mark krijgt alleen afnamen te zien van (groepen van) kinderen waarvoor hij geautoriseerd is. 3 Mark maakt vervolgens gebruik van zoekvelden om de benodigde afnameresultaten te laten verschijnen. Hij vult de

	persoonscode van het kind en drukt daarna op de knop ‘zoeken’. 4 De applicatie toont een nieuwe lijst met resultaten die voldoen aan Mark's zoekcriteria. Op het scherm verschijnen twee afnameresultaten die op verschillende datums zijn afgenomen. 5 Mark vinkt de eerste afname aan. Hij drukt vervolgens op de knop ‘toon resultaat’. 6 Op het scherm verschijnt de uitslag van de afname. Met behulp van staafdiagrammen worden probleem indicerende scores op schalen van de afgenomen vragenlijst getoond. Het eerste diagram geeft de algemene probleemindicatie van het kind weer. Daaronder verschijnen nog drie diagrammen met indicaties van internaliserend en externaliserend gedrag en ADHD. 7 Mark drukt daarna op de knop ‘toon specificatie’. Hij wacht op de specificatie van de afname. 8 Er verschijnt een nieuw scherm met gedetailleerde informatie over de afname. Hierbij worden de gestelde vragen en gegeven antwoorden worden in volgorde van afname getoond.

<<Plaats voor evt. onderschrift - Geen onderschrift? dan deze paragraaf verwijderen! >>

Use cases

Elke scenario is een instantie van een use case, oftewel een use case specificeert alle mogelijke scenario’s voor een gegeven functionaliteit van het systeem. Use cases beschrijven specifieke functies dat het betreffende systeem kan uitvoeren. Een use case wordt geformuleerd door dialogen te modelleren van actoren van het te ontwikkelen systeem.

Use case: StartVragenSessie

Tabel 4.3 UML StartVragenSessie use case

Use case naam	StartVragenSessie
Participerende actoren	Initialisatie door `Cliënt` Communicatie met `VragenafnameApplicatie`
Beginvoorwaarde	De `Cliënt` typt de url van de `CliëntApplicatie` in zijn browser (initialisatie afnamemodule). De `CliëntApplicatie` reageert door het tonen van het inlogscherm (beginscherm) aan de `Cliënt`. Daar vult hij zijn gebruikersnaam en wachtwoord in.
Gebeurtenissen	De `CliëntApplicatie` begint met de controle van de gegevens. Als de gegevenscombinatie onjuist is, vraagt het systeem om de gegevens weer in te voeren. Als de controle slaagt, wordt de `Cliënt` doorverwezen naar de `VragenafnameApplicatie`. De `Cliënt` wordt doorgestuurd naar de startpagina van de `VragenafnameApplicatie`. De `VragenafnameApplicatie` detecteert dat de `Cliënt` toestemming vraagt om vragen te beantwoorden. De `VragenafnameApplicatie` controleert de gegevens in de `ApplicatieDatabase` en haalt extra informatie op met betrekking tot de afname en autorisatie. Na de controle krijgt de `Cliënt` toestemming om te beginnen. De `Cliënt` heeft nu de mogelijkheid om vragen te beantwoorden.

	De sessie wordt beëindigd door de `Cliënt` (als hij drukt op de stopknop) of door de `VragenafnameApplicatie` (als het maximale aantal vragen is bereikt of als er genoeg vragen zijn beantwoord om te kunnen beoordelen met de vereiste betrouwbaarheid)
Afloopvoorwaarde	Voor de `Cliënt` verschijnt de mededeling dat er geen vragen meer gesteld zullen worden. Hij wordt bedankt voor zijn deelname en krijgt instructies over de verder te nemen stappen.

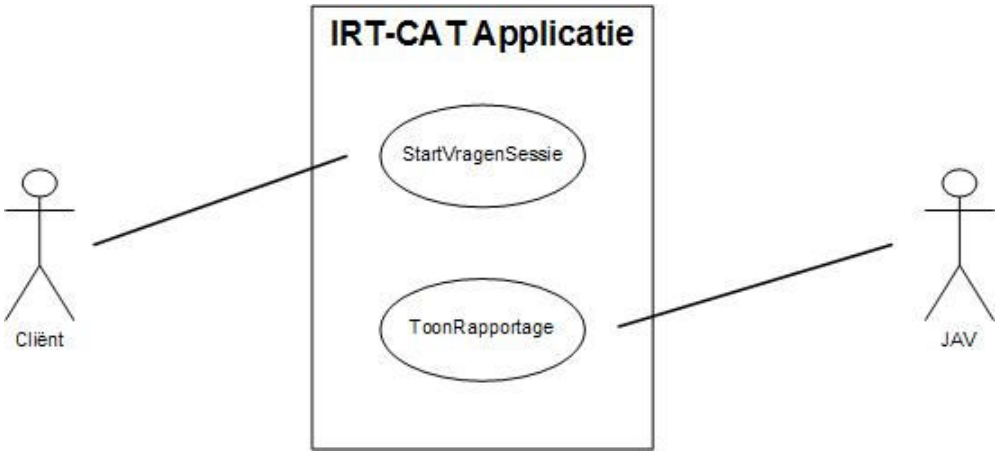
Use case: ToonRapportage

Tabel 4.4 UML ToonRapportage use case

Use case naam	ToonRapportage
Participerende actoren	Initialisatie door JAV
Beginvoorwaarde	1 De JAV typt de url van de `JAV-applicatie` in zijn browser. De `JAV-applicatie` reageert door het tonen van het inlogscherm (beginscherm) aan de JAV. Daar vult hij zijn gebruikersnaam en wachtwoord in.
Gebeurtenissen	2 De `JAV-applicatie` begint met de controle van de gegevens. Als de gegevenscombinatie onjuist is, vraagt het systeem om de gegevens weer in te voeren. Als de controle slaagt, toont de `JAV-applicatie` een scherm met afnameresultaten. De JAV krijgt alleen afnamen te zien van (groepen van) kinderen waarvoor hij geautoriseerd is. 3 De JAV maakt vervolgens gebruik van zoekvelden om de benodigde afnameresultaten te laten verschijnen. Nadat de zoekvelden zijn ingevuld, drukt de JAV op de knop ‘zoeken’. 4 De `JAV-applicatie` toont een nieuwe lijst met resultaten die voldoen aan de zoekcriteria van de JAV. Indien geen enkele afname voldoet aan de criteria, ontvangt de JAV een melding op het scherm 5 Indien er afnamen zijn die voldoen aan de criteria, worden die op het scherm getoond. De JAV vinkt daarna de resultaten aan van de benodigde afname. Hij drukt vervolgens op de knop ‘toon resultaat’. 6 Op het scherm van de `JAV-applicatie` verschijnt de uitslag van de gekozen afname. Met behulp van staafdiagrammen worden probleem indicerende scores op schalen van de afgenomen vragenlijst getoond. Het eerste diagram geeft de algemene probleemindicatie van het kind weer. Daaronder verschijnen nog drie diagrammen met indicaties van internaliserend en externaliserend gedrag en ADHD. 7 De JAV drukt daarna op de knop ‘toon specificatie’. Hierdoor vraagt hij de specificatie van de afname. 8 Er verschijnt een nieuw scherm met gedetailleerde informatie over de afname. Hierbij worden de gestelde vragen en gegeven antwoorden worden in volgorde van afname getoond.
Afloopvoorwaarde	9 De JAV kan de rapportage afsluiten door op de knop ‘stop’ te drukken.

Use case diagram

In een use case diagram worden alle bij het systeem behorende use cases afgebeeld. Het use case diagram bevat alle actoren, use cases (in de vorm van ellipsen), de communicatielijnen tussen actoren en use cases (en eventueel tussen use cases onderling) en de systeemgrens.



Figuur 3 – UML use case diagram van IRT-CAT Applicatie

Objectmodel

Een objectmodel beschrijft de structuur van het systeem in termen van objecten, attributen, associaties en operaties. Een objectmodel wordt weergegeven met een klasse diagram. Klassen zijn abstracties die de gemeenschappelijke structuur en gedrag van een groep objecten specificeren. Objecten zijn in feite niets anders dan instanties van klassen, die gecreëerd, bewerkt en vernietigd worden tijdens systeemuitvoer (executie).

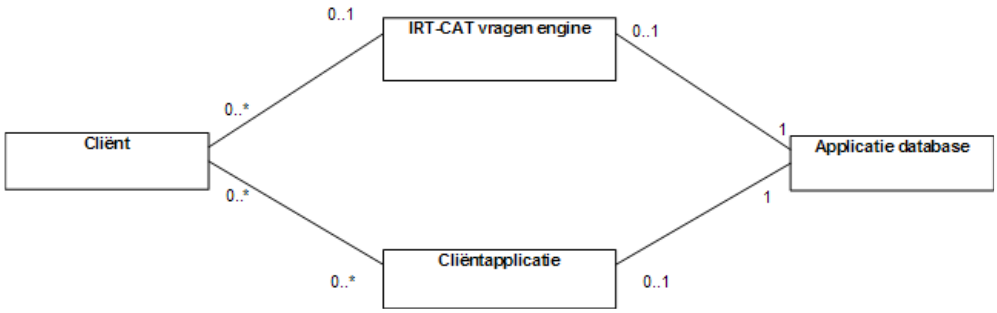
Table 4.5 Objecten van de StartVragenSessie use case

Cliënt	De gebruiker die vragen beantwoord. Een Cliënt meldt zich aan (login) om gebruik te maken van de VragenafnameApplicatie en begint na goedkeuring met een vragensessie. Een Cliënt kan de sessie zelf beëindigen.
CliëntApplicatie	Applicatie die de persoonsgegevens verwerkt. Een Cliënt komt via een Internet url op de inlogpagina van de CliëntApplicatie. De applicatie verwijst de gebruiker aan de hand van de inloggegevens wel of niet door naar de VragenafnameApplicatie.
VragenafnameApplicatie	Maakt gebruik van MR Interview, geconfigureerd met een IRT-CAT script. De VragenafnameApplicatie verwerkt antwoorden van Cliënten en toont na ieder antwoord de volgende vraag. De VragenafnameApplicatie kan de sessie zelf beëindigen.
ApplicatieDatabase	Een SQL Server database, met informatie die tijdens sessies niet door MR Interview wordt opgeslagen. In de ApplicatieDatabase is er onder andere informatie over items, JAVs en Cliënten.

Tabel 4.6 Objecten van de ToonRapportage use case

JAV	De gebruiker die rapportages opvraagt van afnamesessies. Een JAV meldt zich aan (login) om gebruik te maken van de JAVApplicatie. Na goedkeuring kan de JAV informatie opvragen van afnamesessies.
JAVApplicatie	Applicatie die de persoonsgegevens verwerkt en de informatie over afnamesessies vertoont. Na goedkeuring van de inloggegevens kan de applicatie gebruikt worden voor het vertonen van de gestelde vragen, de gegeven antwoorden en de bereikte resultaten (uitslagen).
`ApplicatieDatabase`	Een SQL Server database, met informatie die tijdens sessies niet door MR Interview wordt opgeslagen. In de `ApplicatieDatabase` is er onder andere informatie over items, JAVs en Cliënten.

Klasse diagrammen (class diagrams)

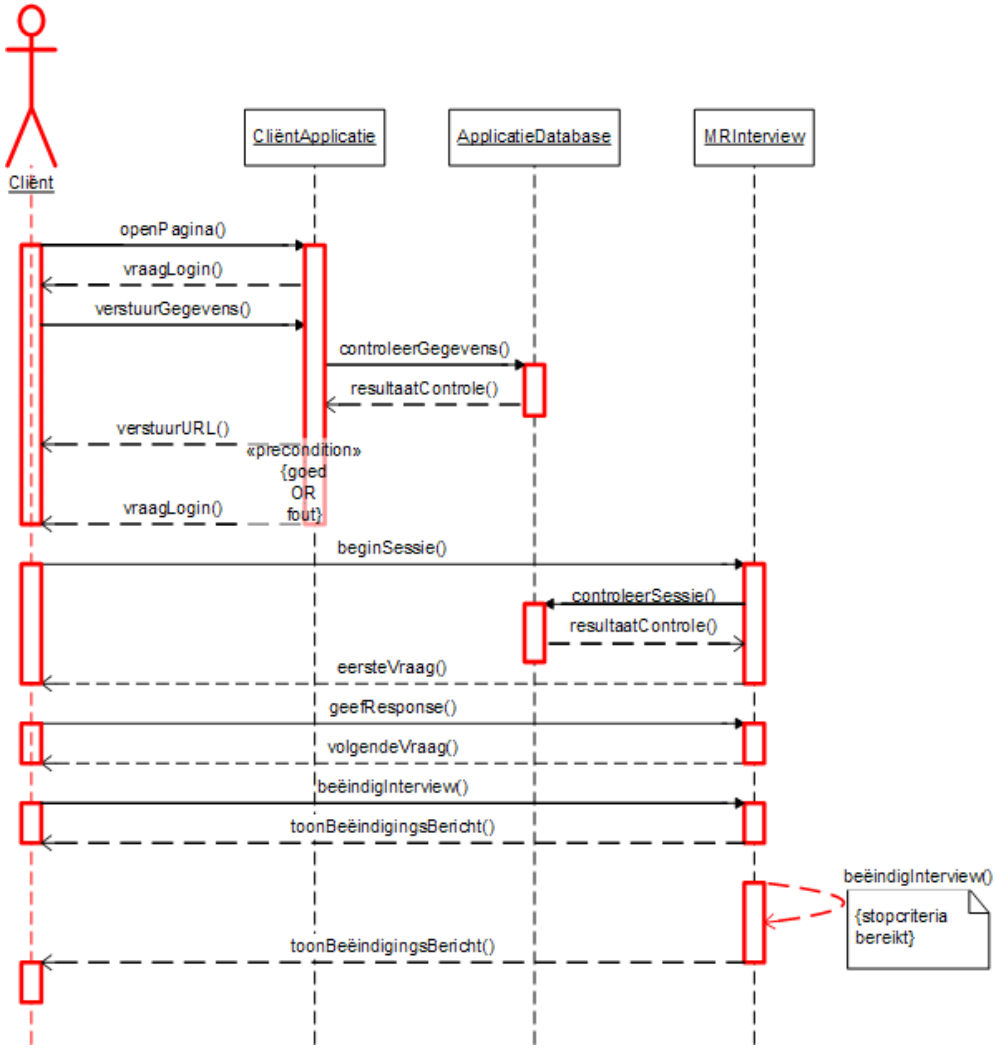


Figuur 4 – UML klasse diagram Client

<JAV klasse diagram volgt nadat datamodel is opgeleverd>

Dynamische modellen

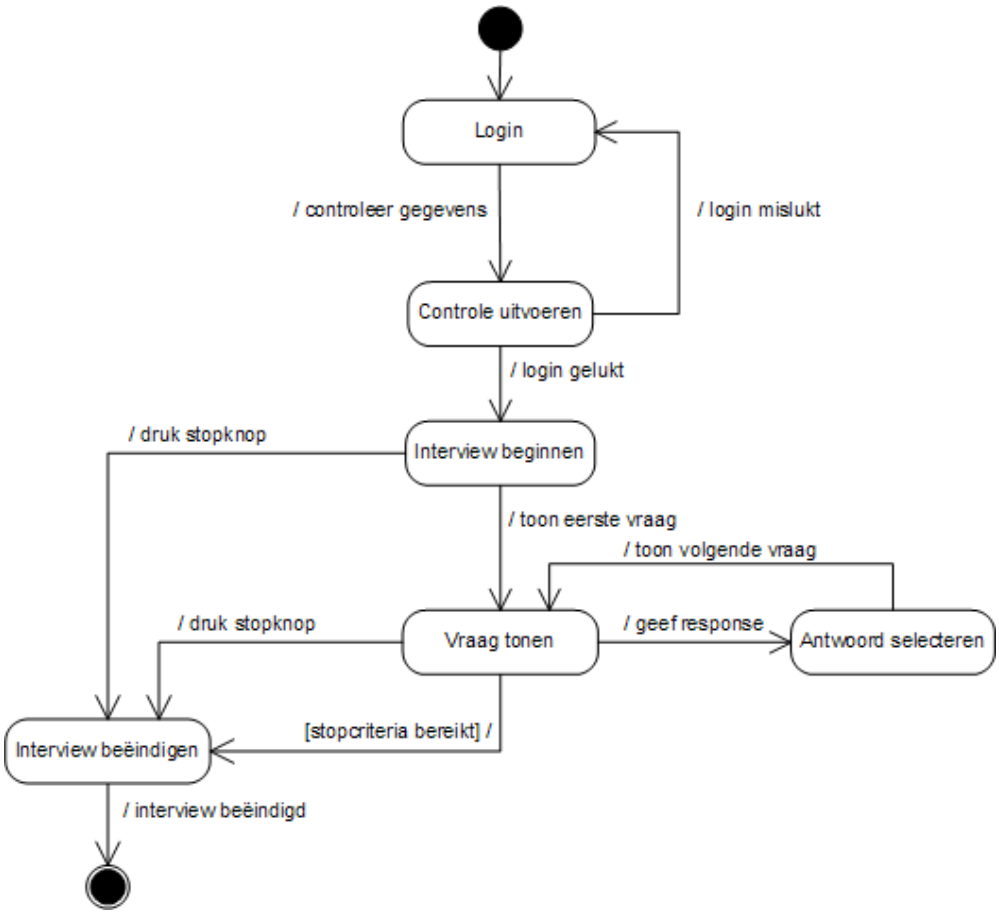
Sequentiediagrammen (sequence diagrams)
Een sequentiediagram verbindt use cases met objecten. Het laat zien hoe de informatie-stroom tussen objecten van een use case (of scenario) verloopt.



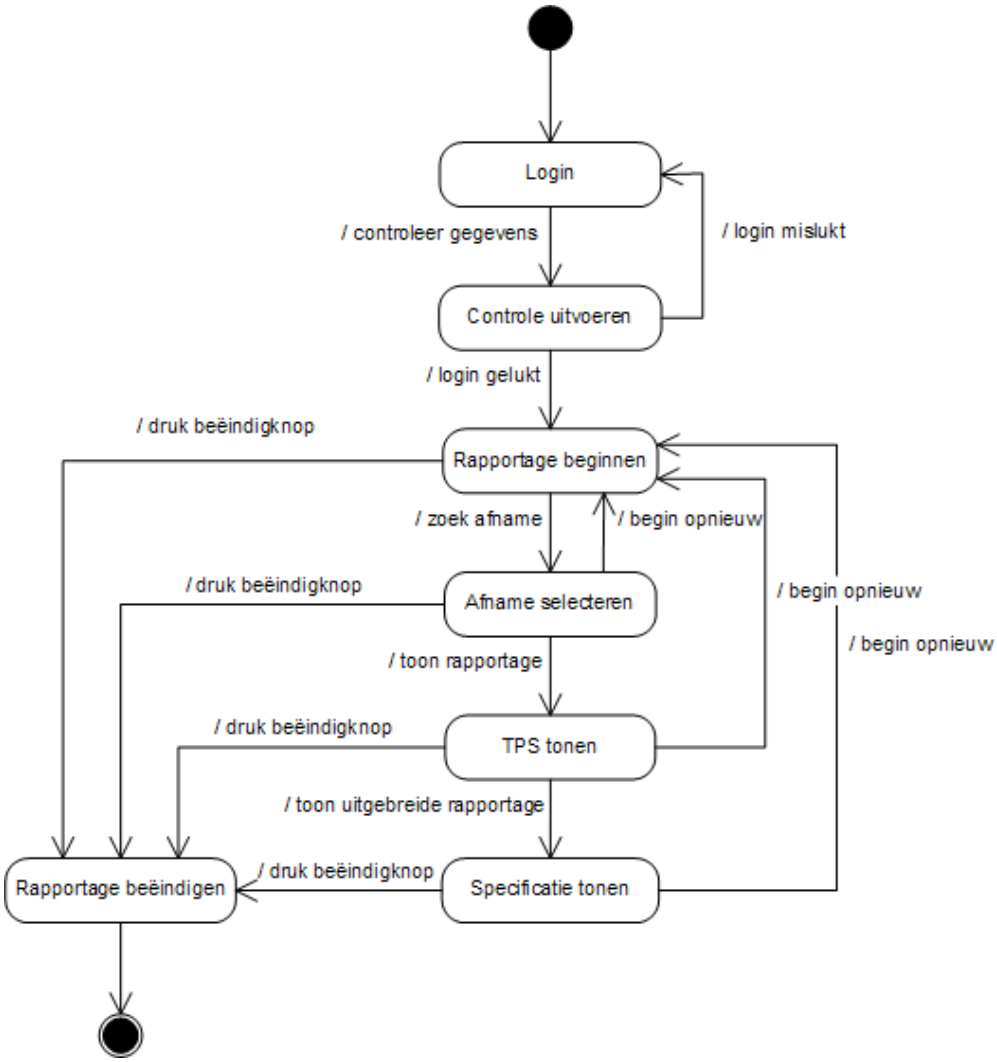
Figuur 5 - Een UML sequentiediagram voor de StartVragenSessie use case.

<ToonRapportage sequentiediagram volgt nadat datamodel is opgeleverd>

Toestanddiagrammen (statechart diagrams)



Figuur 6 - Een UML toestanddiagram voor de StartVragenSessie use case.



Figuur 7 - Een UML toestanddiagram voor de ToonRapportage use case.