TRANSPLANT INTERNATIONAL

Transplant International ISSN 0934-0874

ORIGINAL ARTICLE

A psychometric analysis of the Rotterdam Renal Replacement Knowledge-Test (R3K-T) using item response theory

Sohal Y. Ismail,[1] Lotte Timmerman,[2] Reinier Timman,[1] Annemarie E. Luchtenburg,[2] Peter J. H. Smak Gregoor,[3] Robert W. Nette,[4] René M. A. van den Dorpel,[5] Willij C. Zuidema,[2] Willem Weimar,[2] Emma K. Massey[2] and Jan J. V. Busschbach[1]

1 Section: Medical Psychology and Psychotherapy, The Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands
2 Section: Nephrology & Transplantation, Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands
3 Department of Internal Medicine, Albert Schweitzer Ziekenhuis, Dordrecht, The Netherlands
4 Department of Internal Medicine, Sint Franciscus Gasthuis, Rotterdam, The Netherlands
5 Department of Internal Medicine, Maasstad Ziekenhuis, Rotterdam, The Netherlands

Keywords
end-stage renal disease, health education, knowledge, psychometric, renal transplantation.

Correspondence
Sohal Ismail, Psychiatry-Section Medical Psychology and Psychotherapy, Erasmus MC, Postbox 2040, Room GK-1252, 3000 CA Rotterdam, The Netherlands.
Tel.: +31 68 143 7178;
fax: +31 10 704 4695;
e-mail: s.ismail@erasmusmc.nl

Conflicts of interest
All authors have no other interest to disclose.

Received: 25 April 2013
Revision requested: 14 June 2013
Accepted: 26 August 2013
Published online: 7 October 2013

doi:10.1111/tri.12188

Summary

Knowledge is a prerequisite for promoting well-informed decision-making. Nevertheless, there is no validated and standardized test to assess the level of knowledge among renal patients regarding kidney disease and all treatment options. Therefore, the objective of this study was to investigate the psychometric properties of such a questionnaire for use in research and practice. A 30-item list was validated in four groups: (1) 187 patients on dialysis, (2) 82 patients who were undergoing living donor kidney transplantation the following day, (3) the general population of Dutch residents (n = 515) and (4) North American residents (n = 550). The psychometric properties of the questionnaire were examined using multidimensional item response theory (MIRT). Norm references were also calculated. Five items were found to distort ability estimates (Differential item functioning; DIF). MIRT analyses were subsequently carried out for the remaining 25 items. Almost all items showed good discrimination and difficulty parameters based on the fitted model. Two stable dimensions with 21 items were retrieved for which norm references for the Dutch and North American, dialysis and transplantation groups were calculated. This study resulted in a thorough questionnaire, the Rotterdam renal replacement knowledge-test, which enables reliable testing of patient's knowledge on kidney disease and treatment options in clinic and research.

Introduction

Knowledge can be seen as a prerequisite for promoting well-informed shared decision-making [1,2]. Appropriate knowledge about risks and benefits of different treatment options is necessary to allow well-informed decision-making [3]. Research in predialysis patients shows the benefits of knowledge and awareness on their decision-making, resulting in lower mortality risks, and punctual transplantation access [4,5]. Insufficiency of knowledge plays an even stronger role when patients are considering living kidney donor transplantation (LDKT), because this is more or less a free choice and involves a healthy person [6,7]. For defining the individual knowledge level, an operationalization of 'knowledge' using a questionnaire is needed. However, a validated and standardized test of the level of knowledge among renal patients regarding kidney disease and all treatment options is not available. Existing questionnaires focus either solely on one treatment option or on kidney disease [8–11]. A standardized test of the knowledge regarding kidney disease and treatment options has two clear practical applications. Firstly, to identify possible gaps in patients' knowledge in order to support informed decision-making from a clinical point of view. Secondly, to have a validated instrument which is also sensitive to changes in knowledge. This can then serve the implementation and testing of educational efforts from a research point of view.

Therefore, in this study, we aimed to develop and explore the psychometric properties of a questionnaire on knowledge of renal replacement therapies and kidney disease by applying a multidimensional analysis to validate its use in research and clinic. Additionally, the most optimal methodology for calculating norm references was investigated.

Patients and methods

Phase 1: Development of the Rotterdam renal replacement knowledge-test (R3K-T)

To generate an item pool, we searched the literature for available questionnaires measuring knowledge on renal replacement therapy. This resulted in three measures [9,12,13]. We also consulted members of the Dutch Renal Patient Society and experts in the area of nephrology and transplantation (e.g. nephrologists, psychologist, transplant coordinators, social workers) for additional items that were not represented in the existing measures. This resulted in a pool of 61 items, after deleting duplicate items. To ensure face validity, six patients were asked to comment on these items and give feedback on, for example, the formulation and relevance of the questions and appropriateness of the responses. Based on this information, the items were adapted. Next, a pilot study was conducted among dialysis patients in the Netherlands (n = 116) to investigate the content validity of these items that measured knowledge on kidney disease, dialysis and transplantation options. To further examine validity, we performed an exploratory factor analysis [14]. For most items, the response categories were true/false/don't know. However, considering the nature of several questions (e.g. 4, 12, 22), some items have a multiple choice format with four different categories. In the supplementary material, the questionnaire is included to review the different response options for all the items. Nevertheless, there is always only one correct answer. In order to avoid the possibility of patients responding without knowing the answer, the response option of 'I don't know' was added to all items (scored as if it were an incorrect answer). Thirty items were retrieved from this analysis. We refer the reader to our technical report for a detailed description of the development of this 30-item version as well as the translated versions of the test [15].

Phase 2: Validation of the R3K-T

Participants

The 30-item list is validated in this study in patients on dialysis (n = 187, response rate = 31.3%) and patients who were undergoing a living donor kidney transplantation the next day (n = 82, response rate = 53.9%) in the Netherlands. For labelling purposes, we refer to former group as the 'Dialysis group' and the latter group as the 'Transplant group'. The dialysis patients in this study were recruited from four different dialysis centres, which have approximately a total of 600 patients per year. Patients were asked to participate once they came to the hospital for a dialysis session or for a consultation with their nephrologist. Patients in the transplant group were recruited from the Erasmus Medical Center, Rotterdam, which carries out one-third of the national living donor kidney transplantations per year. The tests were administered in seven different languages. In addition, the test was administered to two representative reference groups, 515 Dutch residents (response rate = 53.0%) and 550 North American residents (response rate = 78.0%) using a Web-based survey. These two reference groups were carefully selected by an independent research bureau to be representative for the general Dutch and the North American populations with quality controls based on age, gender, education, postal code, employment status, attitudes and lifestyle. The research bureau uses an 'open-door' flexible sourcing in which they include anyone who wants to share their honest opinions – even those who would never join a survey panel are recruited through social media, online communities and affiliated partners. The quality of the data is reassured by an external auditing agency (Centre for information-based decision-making & marketing research). The research bureau that performed the sample selection for this study was certified with regards to the required standards regarding the representativeness of the samples. Ethical approval was sought from the Medical Ethical Committee of the Erasmus MC, Rotterdam, but was deemed unnecessary for this kind of noninterventional research.

Analysis

Differential item functioning (DIF) was used to investigate the degree to which some items advantage or disadvantage certain participants groups with respect to the estimates of their knowledge ability using the LTM package in R (2.13.0, R Development Core Team 2011, Rotterdam, the Netherlands) [16]. The rationale of DIF analyses is to identify items that distort the ability estimates for participants and thus jeopardizes correct overall test measurements [17]. Items that are identified to distort test measurement are referred to as having DIF. If a particular item contains DIF, then this is indicated by significant P-values for the DIF parameter. The goal of DIF analyses is to remove items with DIF from the test early on in the validation process.

After removal of items with DIF, the initial test validation process can be undertaken. Classical test theory (CTT) has been the mainstay of psychological test validation from the initial explosion of testing in the 1930s. However, CTT is now being rapidly replaced by item response theory (IRT) as the mainstream basis for educational and psychological test development [18]. A critical assumption of IRT is unidimensionality, which refers to whether patients' responses to an item are accounted for by their 'first' factor score (in terms of CTT) and not by other factors. Thus, unidimensionality was tested in two ways: (1) modified parallel analyses (MPA) incorporated in the ltm package were employed to test for the probability of unidimensionality ($\alpha = 0.05$) [16, 19] and (2) the rule of thumb that the ratio of the first to the second 'eigenvalue' should be above 3 [20]. Accordingly, Rasch, 2PL and 3PL IRT models were fitted using the LTM R package [16]. The assumption of unidimensionality has been relaxed with the development of multidimensional item response theory (MIRT) approach. This was seen as a necessary extension of IRT as most psychological constructs are multidimensional in nature [21]. Therefore, we sought to examine the psychometric properties of the R3K-T using MIRT [22]. MIRT analyses were carried out using the MIRT package for dichotomous data in the R software package [23]. The MIRT analyses return alpha and β-parameters for each item per dimension. The alpha parameter indicates the discriminative power of that item. Items with higher scores on this alpha parameter are better able to discriminate between knowledgeable and less knowledgeable patients. The beta parameters present the difficulty for the imminent dimensions. In MIRT, the interpretation of this parameter is that higher absolute scores indicate easier items: scores towards zero indicate difficult items. Squared β-parameters indicate the degree in which a certain item explains the variance within a certain dimension [24]. The statistical cut-off for the β-parameters is 0.32, because items with this factor loading explain the variance in a certain factor for 10% ($0.32^2 = 0.10$), which is considered to be poor.

We sought to determine the optimal number of dimension for the test using MIRT analyses. Comparisons between different models with varying numbers of dimensions were performed by (1) applying a deviance test (chi-square test) and (2) comparing the differences in Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). Both AIC and BIC use penalty terms for the number of estimated parameters in the different models to prevent the model from overfitting. Overfitting is a statistical problem, which occurs when the fitted model describes noise instead of a true structure. The BIC is from a statistical point of view preferable as it is more able to prevent the model from overfitting than the AIC [25]. Lower AIC and BIC values indicate a better fitting model. For the selection of the appropriate number of dimensions, statistical solutions and content-driven arguments need to be weighted.

Finally, in order to make the R3K-T ready for practical use, test scores were calculated using the CTT approach and compared with test scores generated through IRT analyses. In CTT, the test scores are simply a sum of the number of correctly answered items (each correctly answered item is assigned one point). Thus, we do not take into account that one respondent might have only correctly answered easy items and that another respondent only the difficult ones. On the contrary, IRT scores account for the level of difficulty per item. These scores are presented as the probability that a respondent answers a particular item correctly given the difficulty of that item and the percentage correctly answered items over the total test. Next, the scores generated in terms of CTT and those that were generated via the IRT analysis were compared using a Pearson correlation test. If CTT scores are a close approximation of the IRT-derived test scores, then CTT scores may be favoured for practical proposes as they are easier to calculate. Norm references, cut-offs and reliable change index (RCI) scores were calculated for the total scale and any produced subscales [26]. Cohen's *d* effect sizes are used to estimate the magnitude of the RCIs [27]. Additionally, a score sheet based on percentiles was provided for the total test scores to ease the clinical interpretation of the test scores.

Results

Participant characteristics

Table 1 shows socio-demographic characteristics of the participants. There were significantly more males in the transplantation group ($P = 0.02$) compared with the other groups. The two patient groups also showed a significantly higher proportion of participants with a low educational level ($P < 0.001$). With respect to registration as an organ donor after death, significantly more participants from the Dutch reference group reported to be registered compared with the other three groups ($P < 0.001$). For the patient groups, no differences were found with respect to the distribution of their original disease and duration of dialysis.

Differential Item Functioning

Five items were identified as candidates to be deleted from the list and subsequent analysis due to distorted ability estimates, as indicated by a significant DIF. For example, patients on dialysis would get relatively lower ability estimates on item 6 'How many times a week does a haemodialysis patient generally undergo dialysis?' when compared to the reference groups ($P = 0.002$). This distortion also holds for item 1 ($P < 0.001$), 16 ($P < 0.001$), 23 ($P = 0.002$) and 28 ($P = 0.043$). Therefore, these five items were excluded from further analyses. Items 22 ($P = 0.010$) and 27 ($P = 0.026$) also showed statistically significant DIF

Table 1. Participant characteristics (N = 1335).

Characteristics	United States (N = 550)	Dutch (N = 515)	Dialysis (N = 187)	Transplantation (N = 82)
Gender (male %)	46.5	47.8	46.5	67.9
Mean age in year (SD)	44.8 (16.4)	47.9 (15.5)	54.1 (13.3)	53.58 (13.2)
Educational level (1–3%)*	1 = 3.8	1 = 1.4	1 = 20.9	1 = 15.9
	2 = 57.3	2 = 67.0	2 = 42.6	2 = 57.3
	3 = 38.9	3 = 31.3	3 = 12.2	3 = 20.7
	Missing: 0.0	Missing: 0.3	Missing: 24.4	Missing: 4.9
Mean Duration test in min (SD)	6.9 (3.3)	7.73 (4.0)	NA	NA
Registered for organ donation after death %	39.1	58.8	30.1	31.7
Mean mo on dialysis (SD)	NA	NA	20.4 (22.6)	13.0 (19.3)
Original disease (%)	NA	NA		
Glomerular nephropathy			13.4	12.0
Congenital hereditary			15.9	13.3
Hypertension			15.9	13.3
Diabetes mellitus			17.1	19.3
Systemic diseases			4.9	2.4
All other			32.9	39.8

NA, not available.
Values in the table are presented as *n* with the percentage in parentheses or mean values with ±SD in parentheses.
*The educational level was valued at three levels; 1 = low; elementary school, 2 = average; high school and 3 = high; college degree (+some post-graduate/professional degree).

parameters, however, the direction of the ability estimates is consistent with the expected knowledge level of the respective groups: these items favour the knowledgeable groups more than the reference groups. For item 22, the proportion of the North American, Dutch, Dialysis and Transplant participants with a correct answer was 0.57, 0.70, 0.70 and 0.87, respectively: for item 27, the respective proportions were 0.30, 0.17, 0.60 and 0.80. The consistency in the estimation is reflected in the increasing nature of the proportions as the expected knowledge level of the different groups increases.

Multidimensional item response theory

Before fitting an appropriate IRT model, the critical assumption of unidimensionality was assessed. The test for unidimensionality using MPA on all participants together was significant (P = 0.009). This indicates multiple dimensions. The ratio of the first to the second eigenvalue was 7.342/2.191 = 3.35. Given these findings in support of multidimensionality, MIRT was performed on the remaining 25 items. A single-dimensional model, postulating general knowledge was tested against a two-, three- and four-dimensional model. Table 2 shows that the difference between the models up to model 3 is significant at α = 0.05 level. The decreases in the log-likelihood and the AIC also show that the additionally estimated parameters for the third dimension could provide further significant information. Table 3 displays the parameters for the items in the third factor. The best factor loading in that third factor explains 24% ($R^2 = 0.491^2 = 0.241$) of the variance in that factor: this explained variance is considered to be fair. From a conceptual point of view, we judge that this third factor contains items which cover varying subject areas [insurance (5), kidney disease (8), treatment option (17) and immunology (21)]. Also the BIC, which is for statistical reasons preferred over the AIC, increased when moving to three-dimensional model. In other words, the two-dimensional

Table 2. Comparing multidimensional item response theory models.

Model	Log-likelihood	AIC	BIC	Comparing models
1	−15072.57	30229.14	30447.33	
2	−14679.19	29484.39	29811.68	(mod1 versus mod2; χ^2 = 786.75, d.f. = 20, P < 0.001)
3	−14610.98	29389.96	29826.35	(mod2 versus mod3; χ^2 = 136.43, d.f. = 19, P < 0.001)
4	−14648.25	29506.50	30052.00	(mod3 versus mod4; χ^2 = −74.54, d.f. = 18, P = 1.000)

This table shows the log-likelihood, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) parameters for the fitted models with the mirt() function. The final column shows the comparisons between the nested models using a deviance test (chi-square statistic, degrees of freedom, *P*-value). The models reflect the number of dimensions tested ('mod1' contains one dimension, 'mod2' contains two dimensions, etc.).

Table 3. Third factor items following the factor parameterization in multidimensional item response theory.

Items	Short item description	α (SE)	β_1 (SE)	β_2 (SE)	β_3 (SE)
05	Hospital costs of a living kidney donor transplantation are paid for by the health insurance of the recipient	0.294 (0.092)	0.183 (0.111)	−0.303 (0.087)	0.410 (0.063)
08	Patients with renal disease should not eat too much salty food	0.321 (0.081)	0.136 (0.111)	−0.248 (0.097)	0.491 (0.126)
23	For the treatment of end-stage renal disease, kidney transplantation is preferred over dialysis	0.409 (0.089)	0.370 (0.074)	−0.367 (0.098)	0.371 (0.094)
21	A full match between the tissue of the donor and the patient provides the best survival	0.336 (0.163)	0.142 (0.122)	−0.366 (0.084)	0.426 (0.082)

Items are paraphrased for brevity. This table displays slopes transformed into varimax-rotated factor loadings metric of the deleted items from the third factor: item discrimination parameter [α (SE)] and item difficulty parameters for the three respective dimensions [β*1* (SE), β*2* (SE) *and* β*3* (SE)].

model has a better fit and a more coherent content compared with the three-dimensional model. Hence, we rejected the three-dimensional solution and items 3, 8, 23 and 21 were excluded from further analyses. However, item 5 is viewed by the authors as clinically quite informative. Considering the conceptual nature of this item, it fits the second factor better than the first and it still explains the variance in the second factor fairly (15%; $R^2 = 0.393^2 = 0.154$). Finally, Table 4 presents the factor loadings that were extracted from the two-dimensional solution, producing an easily interpretable and cohesive structures for the 21 items.

After the aforementioned analyses, a two-dimensional R3K-test was constructed based on 21 items. Table 4 presents the two-dimensional solution with the factor labels. The factor labels ‘Dialysis & Transplantation’ (DT) and ‘Living Donation’ (LD) reflect the content of the respective factors and were agreed upon through a consensus meeting

Table 4. Item characteristics on the two subscales of the 21-item R3K-T.

Items	Short item description	α (SE)	β_1 (SE)	β_2 (SE)
Dialysis and transplantation				
04	Peritoneal dialysis: Which part of the body makes this possible?	0.788 (0.137)	0.862 (0.195)	−0.211 (0.129)
12	An advantage of peritoneal dialysis is…	0.758 (0.174)	0.848 (0.180)	−0.197 (0.125)
18	Peritoneal dialysis: What happens with the fluid after it is brought into the abdominal cavity through a catheter?	0.646 (0.125)	0.803 (0.126)	−0.015 (0.113)
24	Is peritonitis one of the biggest problems with patients with peritoneal dialysis?	0.623 (0.087)	0.748 (0.126)	−0.251 (0.087)
26	Certain vitamins are lost during dialysis. If you are on dialysis you are therefore prescribed extra vitamins	0.504 (0.074)	0.687 (0.106)	−0.180 (0.084)
20	How many hours a day is a haemodialysis patient connected to the machine?	0.559 (0.074)	0.650 (0.112)	−0.370 (0.075)
10	Renal replacement therapy is necessary if kidney function is only 50%	0.531 (0.102)	0.643 (0.125)	−0.344 (0.082)
14	A permanent access to the bloodstream is needed for haemodialysis	0.424 (0.065)	0.583 (0.089)	−0.288 (0.068)
27	Kidneys from living donors have longer graft survival than from deceased donors	0.381 (0.078)	0.535 (0.094)	−0.308 (0.071)
17	ESRD: Kidney transplantation is generally preferred to dialysis	0.497 (0.069)	0.514 (0.102)	−0.483 (0.069)
22	Immunosuppressive drugs are administered to transplant patients for	0.325 (0.068)	0.303 (0.080)	−0.483 (0.067)
Living donation				
15	Surgical complications after donation are common in living kidney donors	0.755 (0.086)	0.210 (0.147)	−0.843 (0.114)
11	Donating a kidney increases the risk of developing a kidney disease	0.640 (0.083)	0.172 (0.114)	−0.781 (0.098)
13	Most living kidney donors remain in the hospital for 2 weeks after surgery	0.544 (0.088)	0.334 (0.113)	−0.658 (0.081)
07	Very few living kidney donors have long-term health problems after donation	0.399 (0.072)	0.143 (0.088)	−0.615 (0.079)
25	Kidney donation may affect a woman’s chances of getting pregnant	0.428 (0.076)	0.274 (0.093)	−0.594 (0.093)
19	Most living kidney donors can participate in sports and work within 4–6 weeks after donation	0.446 (0.067)	0.346 (0.092)	−0.571 (0.070)
30	When the kidney of a living donor does not match the recipient, living donation is not an option with this donor	0.383 (0.075)	0.386 (0.091)	−0.484 (0.069)
09	A living kidney donor has to be younger than 50 years old	0.405 (0.068)	0.428 (0.092)	−0.471 (0.067)
02	Only direct family members (brothers, sisters, parents or children) can donate a living kidney	0.201 (0.093)	0.047 (0.082)	−0.446 (0.083)
05	Hospital costs of a living kidney donor transplantation are paid for by the recipient’s health insurance	0.230 (0.071)	0.274 (0.079)	−0.393 (0.067)

Items are paraphrased for brevity. This table displays slopes transformed into varimax-rotated factor loadings metric: item discrimination parameter [α (SE)] and item difficulty parameters for the two respective dimensions [β*1* (SE) and β*2* (SE)].

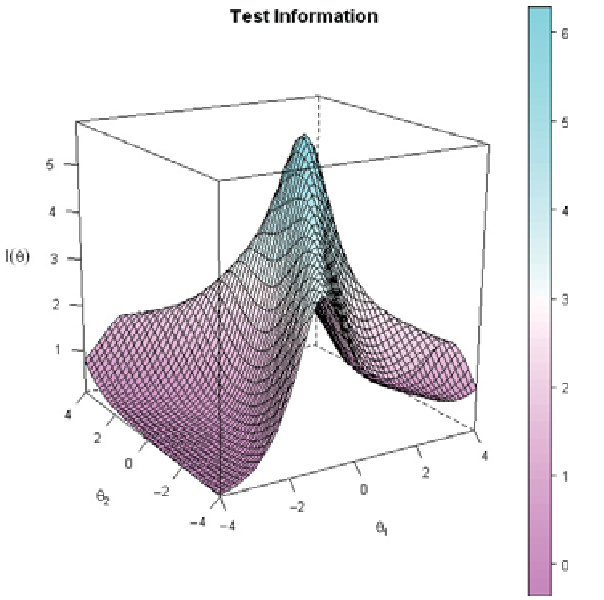

Figure 1 This figure presents the two-factors test information plot. On the *x*-axis, the ability levels are shown, and on the *y*-axis, the discriminative power of the tests. The 21 items are represented under the surface of the 'cloth'. The peak in the middle of the figure shows that both factors have the best discriminative power for patients with an average ability level (0). As one moves towards the end points of the ability scale (−4 and 4), the figure shows that the discriminative power decreases.

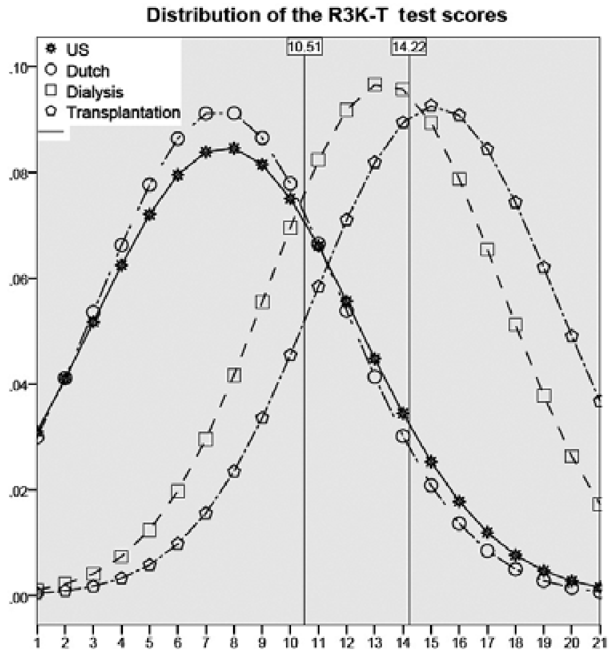

Figure 2 This figure shows the parametric distribution of the test scores with the accompanying cut-off values between the Dutch versus dialysis and dialysis versus transplantation groups. The cut-off values are represented as interpolation lines in this figure. The *x*-axis are the test scores displayed (min = 0, max = 21) and on the *y*-axis the respective proportion of participants with that score.

with the authors. A didactic presentation of these two factors is presented in Fig. 1. The first factor (θ_1) retrieves the most test information at moderate knowledge levels, whereas the second factor (θ_2) contains items that retrieve test information across almost the whole range of knowledge levels.

Norm-referenced test scores and RCIs

The CTT total sum score method was compared with the IRT ability estimates. The correlation between these two approaches on the total individual scores was very high ($r = 0.986$, $P = 0.001$). Table 5 presents the normative data for interpreting the test scores based on CTT scores for the total and subscale scores. There were no significant differences in the total group scores between the Dutch and North American general population groups ($P = 0.955$). The more knowledgeable group – dialysis patients – showed significantly higher test scores compared with their reference group (Dutch; $P < 0.001$). The

Table 5. Norm-reference scores and Reliable Change Index (RCI).

Norm-references R3K-T	US (*N* = 550)	Dutch (*N* = 515)	Dialysis (*N* = 187)	Transplantation (*N* = 82)
R3K-T total score (SD)	7.68 (4.71)	7.51 (4.35)	13.35 (4.12)	15.14 (4.31)
RCI (Cohen's *d*)			4.30 (1.36)	5.42 (0.43)
Cut-off (*P*-value)			10.51 (<0.001)	14.22 (0.002)
Subscale DT (SD)	3.57 (2.66)	3.32 (2.32)	7.91 (2.29)	8.35 (2.54)
RCI (Cohen's *d*)			2.64 (−1.98)	3.59 (0.18)
Cut-off (*P*-value)			5.63 (<0.001)	8.12 (0.023)
Subscale LD (SD)	4.11 (2.65)	4.19 (2.58)	5.43 (2.49)	6.89 (2.48)
RCI (Cohen's *d*)			2.92 (−0.48)	3.77 (0.58)
Cut-off (*P*-value)			4.82 (0.003)	6.16 (<0.001)

This table presents mean group scores (standard deviation), cut-off scores (significance level, $\alpha = 0.05$) and RCI scores (Cohen's *d* effect size) for the total test scores and the subscale scores. The cut-off scores present the cut-off value between that group and the preceding. The RCI scores present the minimum improvement a person in that group should make for a reliable shift to the next group. The subscale DT, dialysis and transplantation and subscale LD, living donation.

Table 6. Score sheet of the 21-item R3K-T based on percentiles.

Percentiles %	10	20	30	40	50	60	70	80	90
US	0–1	2–3	4	5–6	7–8	9	10–11	12	13–21
Dutch	0–2	3	4–5	6	7	8–9	10	11	12–21
Dialysis	0–8	9–10	11–12	13	14	15	16	17	18–21
Transplantation	0–8	9–11	12–14	15	16	17	18	19	20–21

This table presents the score sheet for the knowledge test for each group separately. The percentiles indicate the total test score below which a certain percent of the participants fall. For example, the 40th percentile is the score below which 40% of the participants may be found. The test scores vary from 0 to 21.

'most' knowledgeable group in this study – transplant group – showed superior test scores compared with the dialysis patients ($P = 0.016$). The optimal cut-off scores between the Dutch, dialysis and transplant group (respectively 10.51 and 14.22) represent significant shifts to a better-informed group. Figure 2 graphically displays the distribution of the test scores for the different groups and the two significant cut-off scores. The RCIs show that the clinical shifts between the Dutch dialysis and dialysis–transplant groups have large and medium effect sizes for the total scale and the two subscales (Table 5). Finally, Table 6 provides the percentiles for the four study groups. These percentiles indicate the total test score below, which a certain percentage of the participants fall. For example, the 40th percentile is the score below, which 40% of the participants may be found.

Discussion

This study aimed to develop and validate an instrument with which knowledge of kidney disease and the related treatment options can be reliably measured. As knowledge measurement is indirect, a patient's ability must be inferred from test scores. In this validation study, we used a multilevel approach to infer these test scores. Most items showed good discrimination parameters based on the model that was fitted, which indicate that patients with various knowledge levels can be discriminated adequately.

The result of this study is a knowledge test with two solid dimensions. The deviance test and the AIC indicated that there was potentially a third dimension, while the BIC points towards a two-dimensional solution. When we looked for content-driven argumentation for a potential three-dimensional solution, we found conclude that the third dimension is incoherent and was thus not of additional value for the practical use of the test. Hence, we conclude that a two-dimensional test is the best solution in this case. All the items in the two scales show fair-to-excellent item properties. Only item 22 about immunosuppressive drugs had psychometric shortcomings. However, this item was regarded as conceptually unique and informative. One practical solution for improving the test performance of item 22 may be the reduction in the number of response categories: dichotomizing the response categories into a true/false format [28].

For the practical use of the test in the clinic and research, the most optimal scoring method was determined. Determining test scores for the 21 items using both CTT and IRT, showed very high correlation. In the light of this, the CTT method is preferable because these calculations are easily performed by hand: each correct response receives a score of 1 and an incorrect response a score of 0. From a clinical point of view, if patients are to make a well-informed decision regarding renal replacement therapy, they should score highly on both subscales as this indicates sufficient knowledge of all their treatment options. Therefore, total scores are preferred. Subscale scores become relevant when one identifies a certain group for which it has clear clinical implications to contain knowledge on a specific scale. For instance, for potential donors, it may be relevant to have sufficient knowledge on the 'Living Donation' subscale relative to the 'Dialysis & Transplantation' subscale. Further validation of the R3K-T in more specific groups, such as living kidney donors, for which subscale scores may contain clinically relevant information would increase practical rigour of this test. Furthermore, as most people are more familiar with percentiles rather than RCI scores, a score sheet based on percentiles is also presented to ease the practical use of the test.

Several limitation of this study may be considered: firstly, a validation of this test in the population of predialysis patients is unfortunately lacking in this study. This subpopulation is very interesting because they are yet to the make a (well-informed) decision regarding the different renal replacement therapies. Secondly, given the sample size of the two clinical groups in this study, representativeness of these populations can be questioned. However, given the relatively low number of items with DIF and the relative ease at which the analyses returned a valid and cohesive two-dimensional structure, the sample size and the distinctiveness of the different subgroups did not indicate any flaws regarding the samples for now. Nonetheless, future validation studies using the flexibility of IRT analyses to easily incorporate additional and larger samples are warranted. This is particularly needed for the predialysis group but also for the dialysis group given the relatively low response rate. Thirdly, patients interested in receiving a LDKT who were ruled out and patients who are on the waitlist for a deceased donor kidney transplantation but who are not willing to get a LDKT may be clinically distinctive from the general dialysis population. These potential subgroups were not treated as a separate group in this study and categorized according to their current treatment (dialysis). The design for the current study was set out to make an initial distinction between the most apparent groups, namely the general population, dialysis patients and transplant patients. Nevertheless, we appreciate that from a clinical point of view, more subpopulations may exist for which statistical validation is needed. However, with current data, no sufficient variation is present to make such a distinction statistically. Fourthly, the items on financial matters were rather weak in their test performance when administered across borders. These financial items were valid in the Dutch test case [15]. However, in this initial cross-cultural validation using data from North American residents, the items on financial issues dropped dramatically in their test performance. This is likely to be due to different legislations when it comes to reimbursement of health-related costs by public funds or insurance companies. Yet, item 5 ‘Hospital costs of a LDKT are paid for by the health insurance of the recipient’ remained in the questionnaire. It is psychometrically not a very strong item, but it could be regarded as a starting point from which one or two extra items on financial issues could be added in line with the local legislation. This point on the financial items could be seen as a limitation. However, given the ever changing nature of insurance legislation and policy on financial compensation, it will be difficult to make standardized items that can be used irrespective of time or setting.

In conclusion, this is a psychometrically solid, brief and comprehensive self-report test on knowledge of renal replacement therapies that can be implemented in research and the clinic. Further validation of the R3K-T is warranted in larger local subgroups, predialysis patients and in general populations other than those of the United States and the Netherlands. The various translations of the R3K-T could facilitate these international validation studies.

Authorship

WW, JJVB, EKM, SYI, LT, AEL and WCZ: participated in research design. SYI, LT, RT, EKM, PJHSG, JJVB and WW: participated in the writing of the paper. SYI, AEL, LT, PJHSG, RWN and RMAD: participated in the performance of the research. SYI, RT, JJVB and WW: participated in data analysis.

Funding

This study is part of research funded by The Netherlands Kidney Foundation.

Acknowledgements

The authors would like to thank Ton van Kooij, Dianne van Dongen, Ellen Versluijs, Anita Mathoera, Adilson da Silva, Ria Pennewaard, Peter Menor, Liduine Schönau, Anja van Giels and Sonja Faber for assisting in the data collection process.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Enclosed is the English version of the final 21-item questionnaire.

References

1. Rodrigue JR, Pavlakis M, Danovitch GM, *et al.* Evaluating living kidney donors: relationship types, psychosocial criteria, and consent processes at US transplant programs. *Am J Transplant* 2007; **7**: 2326.
2. Morton RL, Howard K, Webster AC, Snelling P. Patient information about options for treatment: methods of a national audit of information provision in chronic kidney disease. *Nephrology (Carlton)* 2010; **15**: 649.
3. Vamos EP, Novak M, Mucsi I. Non-medical factors influencing access to renal transplantation. *Int Urol Nephrol* 2009; **41**: 607. Epub 2009 Apr 2007.
4. Mehrotra R, Marsh D, Vonesh E, Peters V, Nissenson A. Patient education and access of ESRD patients to renal replacement therapies beyond in-center hemodialysis. *Kidney Int* 2005; **68**: 378.
5. Mehrotra R. Bridging the care gap around dialysis initiation: is CKD education part of the solution? *Am J Kidney Dis* 2011; **58**: 160.
6. Zimmerman D, Albert S, Llewellyn-Thomas H, Hawker GA. The influence of socio-demographic factors, treatment perceptions and attitudes to living donation on willingness to consider living kidney donor among kidney transplant candidates. *Nephrol Dial Transplant* 2006; **21**: 2569. Epub 2006 May 2515.
7. Dew MA, Jacobs CL, Jowsey SG, Hanto R, Miller C, Delmonico FL. Guidelines for the psychosocial evaluation of living unrelated kidney donors in the United States. *Am J Transplant* 2007; **7**: 1047. Epub 2007 Mar 1012.
8. Stothers L, Gourlay WA, Liu L. Attitudes and predictive factors for live kidney donation: a comparison of live kidney donors versus nondonors. *Kidney Int* 2005; **67**: 1105.

9. Devins GM, Binik YM, Mandin H, *et al.* The Kidney Disease Questionnaire: a test for measuring patient knowledge about end-stage renal disease. *J Clin Epidemiol* 1990; **43**: 297.
10. Urstad KH, Andersen MH, Oyen O, Moum T, Wahl AK. Patients' level of knowledge measured five days after kidney transplantation. *Clin Transplant* 2011; **25**: 646.
11. Wright JA, Wallston KA, Elasy TA, Ikizler TA, Cavanaugh KL. Development and results of a kidney disease knowledge survey given to patients with CKD. *Am J Kidney Dis* 2011; **57**: 387.
12. Stothers L. Reliability, validity, and gender differences in the quality of life index of the SEAPI-QMM incontinence classification system. *Neurourol Urodyn* 2004; **23**: 223.
13. Rodrigue JR, Cornell DL, Lin JK, Kaplan B, Howard RJ. Increasing live donor kidney transplantation: a randomized controlled trial of a home-based educational intervention. *Am J Transplant* 2007; **7**: 394.
14. Thompson B. *Exploratory and Confirmatory Factor Analysis*. Washington, DC: American Psychological Association, 2004.
15. Ismail SY, Massey EK, Luchtenburg AE, *et al.* Development of the Rotterdam Renal Knowledge-Test (R3K-T). 2011; http://repub.eur.nl/res/pub/23968/Manuscript_Development%20of_R3K-T.pdf.
16. Rizopoulos D. ltm: an R package for latent variable modeling and item response theory analysis. *Appl Psychol Meas* 2006; **15**: 361.
17. Magis D, Beland S, Tuerlinckx F, De Boeck P. A general framework and an R package for the detection of dichotomous differential item functioning. *Behav Res Methods* 2010; **42**: 847.
18. Embretson SE, Reise S. *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum Publishers, 2000.
19. Drasgow F, Lissak RI. Modified parallel analysis: a procedure for examining the latent dimensionality of dichotomously scored item responses. *J Appl Psychol* 1983; **68**: 363.
20. Morizot JM, Ainsworth AT, Reise SP. Towards modern psychometrics: application of item response theory models in personality research. *Handbook of Research Methods in Personality Psychology*. New York: Guilford Press, 2007: 407–423.
21. Wirth RJ, Edwards MC. Item factor analysis: current approaches and future directions. *Psychol Methods* 2007; **12**: 58.
22. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. *J Stat Softw* 2012; **48**: 1.
23. Gibbons RD, Hedeker DR. Full-information item Bi-factor analysis. *Psychometrika* 1992; **57**: 423.
24. Comrey AL. *A First Course in Factor Analysis*. New York: Academic Press, 1973: 640 pp.
25. Schwarz GE. Estimating the dimension of a model. *Ann Stat* 1978; **6**: 461.
26. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; **59**: 12.
27. Cohen J. A power primer. *Psychol Bull* 1992; **112**: 155.
28. Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol* 2009; **5**: 27.

