In recent years, an increasing need to improve rehabilitation care and the outcomes of therapeutic interventions has led to a growing interest in the use of standardized self-reported outcome measures. In the field of orthotics and prosthetics, self-report outcome measures evaluating patients satisfaction with orthosis (PSwO) allow clinicians to evaluate the functional outcome, monitor the quality of orthotics, and improve decision-making. In fact, patient perspectives are an important component of evidence-based practice and play a key role during the course of treatment given that patient satisfaction, which is related to the quality of care, is in turn related to compliance, i.e., to the treatment efficacy versus abandonment rate.

Despite the availability of several outcome measures for PSwO, the majority have been developed in the English language and, to our knowledge, no Arabic version of outcome measures assessing PSwO is currently available. Thus there is a need to produce and validate versions of the best existing outcome measures adapted to the Arabic culture, in particular, considering the growing interest in rehabilitation and occupational therapy in Saudi Arabia.

One of the established measures in the orthotics field is the Orthotics and Prosthetics Users’ Survey (OPUS), which assesses functional outcomes and patient satisfaction in both orthotics and prosthetics users. The questionnaire consists of 5 independent modules that can be completed separately. In the present paper, we focus on the Client Satisfaction with Device (CSD) module, which has recently been modified to create a psychometrically more robust scale. The CSD...
was produced in the US English and has been validated also in the Swedish language.8,10
The aim of this study was to perform a translation and cross-cultural adaptation into Arabic of the revised version of the CSD (Arabian version of the CSD [CSD-Ar]) and analyze its psychometric properties through a modern psychometric approach involving factor analysis and Rasch analysis (RA) to enhance confidence in its use in different clinical settings.

PATIENTS AND METHODS

Subjects
A convenience sample of 100 subjects was recruited between February and June 2013. Subjects were either outpatients or inpatients consecutively referred to the rehabilitation department of two hospitals in Riyadh, Saudi Arabia. Inclusion criteria were: present use of an orthosis as part of the rehabilitation program and age 18 years or older. Exclusion criteria were: problems with understanding Arabic and any diagnosed cognitive deficit.

Clinical characteristics of the sample are reported in Table 1; type of orthosis used and duration of use are shown in Table 2.

A sample size of 100 was selected because in RA, this number is sufficient (for a reasonably targeted sample) to obtain stable calibration of items within ±½ logit with 95% confidence.11

Local ethics committee approval for the study was obtained from the two hospitals. Subjects signed an informed consent provided with the questionnaires. This study was conducted in compliance with the scientific principles governing clinical research as set out in the Declaration of Helsinki.

Instrument—The CSD module
The CSD is a module to assess patients’ satisfaction with their orthosis/prosthesis. The CSD questions address various aspects of the orthosis/prosthesis (e.g., weight, dimensions, and ease of use). The modified CSD version consists of 8 items and is rated on a 4-level Likert scale ranging from 1 "strongly agree" to 4 "strongly disagree."5,8 The translation and cultural adaptation of the CSD-Ar was carried out in accordance with international guidelines,7 following a process that included pilot testing (with cognitive debriefing) and expert analysis, without any major problems being found (Appendix 1).

Translation process and cultural adaptation
The translation and cross-cultural adaptation of the

| Table 1. Characteristics of the sample (N=100). |
|------------------|------------------|
| **Age** | **Years** |
| Mean (SD) | 36.7 (1.6) |
| **Gender** | (%) |
| Males | 59 |
| Females | 41 |
| **Impairment** | (%) |
| Upper limb injury | 69 |
| Lower limb injury | 18 |
| Trunk | 13 |
| **Education** | (%) |
| Illiterate | 8 |
| Elementary | 4 |
| Secondary | 7 |
| High school | 38 |
| Undergraduate | 35 |
| Graduate | 8 |
| **Job** | (%) |
| Student | 17 |
| Soldier | 15 |
| Governmental sector | 28 |
| Private sector | 4 |
| Housewife | 21 |
| Retiree | 9 |
| Unemployed | 6 |

| Table 2. Characteristics of the orthoses (n= 100). |
|------------------|------------------|
| **Type of orthosis** | (%) |
| Hand orthosis | 45 |
| Finger orthosis | 20 |
| Arm orthosis | 5 |
| Shoulder orthosis | 5 |
| Trunk orthosis | 1 |
| Cervical orthosis | 3 |
| Knee orthosis | 3 |
| Knee-ankle-foot orthosis | 2 |
| Ankle-foot orthosis | 10 |
| Unspecified lower limb orthosis | 6 |
| **Duration of use** | (%) |
| Less than a wk | 12 |
| 1-2 wk | 18 |
| 2-4 wk | 30 |
| 1-3 mo | 25 |
| 3-6 mo | 4 |
| More than 6 mo | 11 |
CSD-Ar was carried out in accordance with international guidelines. First, the original English version was translated into Arabic by a professional translator, who had no grasp of the CSD concept, and a bilingual occupational therapist who was acquainted with the CSD. Subsequently, a consensus Arabic version was created by a committee composed of 5 bilingual professionals, expert in these methodological and clinical areas. Then, a native English speaker and a professional translator with no medical background completed the backward translation of the CSD-Ar. After that, the expert committee reviewed and consolidated the two versions of the questionnaire into the prefinal version of the CSD-Ar. This version was tested on 10 subjects, who subsequently underwent a cognitive debriefing, i.e., an interview about the clarity, intelligibility, appropriateness, and cultural relevance of the target language version. Since there were no concerns with the prefinal version testing, this version was accepted as the final version of the CSD-Ar. A copy of the questionnaire can be obtained from the corresponding author.

Data collection
The CSD-Ar was self-administered by the patients under the supervision of a therapist; in the eight illiterate subjects, the questionnaire was administered by the therapist.

Statistical Analysis

Factor Analysis
To fulfill the requirement of RA regarding unidimensionality, and due to the unknown structure of the 8-item CSD-Ar responses, an exploratory factor analysis for ordinal data was conducted prior to proceeding with RA. After checking for the adequacy of the polychoric correlation matrix with the Bartlett and Kaiser-Meyer-Olkin tests, Horn parallel analysis (PA) was performed to estimate the number of meaningful dimensions in the response matrix by comparing the size of eigenvalues obtained from the principal component analysis with those obtained from a randomly generated data set of the same size and number of variables. In addition, the model suggested by PA was evaluated with exploratory factor analysis (EFA), using the unweighted least squares factor extraction method to investigate the contribution of each item to the scale. PA and EFA were conducted using FACTOR 8.1 software (Departament de Psicologia Universitat Rovira i Virgili Tarragona, Spain).

Rasch Analysis
RA is a model-driven process; data collected are compared to an ideal measurement model to see if it behaves in the way that the model predicts. If it does not, statistical information is provided to guide the revision of the measurement instrument so that it functions more effectively as a measurement tool. In recent years, RA has been recommended as a method for assessing the metric properties of rating scales as it provides psychometric information that cannot be obtained with the classical test theory approach. A comparison between RA and CTT is beyond the scope of this paper; however, for a summary of key differences between RA and CTT approaches to developing assessment measures, see work by McAllister (Table 2).

WINSTEPS software version 3.68.2 was used to conduct RA utilizing a rating scale model for the investigations as follows:

1) Rating scale diagnostics of the CDS-Ar was evaluated using the guideline recommendations of Linacre: (a) at least 10 observations per category; (b) even distribution of category use; (c) category outfit mean square values >2; (d) monotonic increase in both average measures of persons with a given category and thresholds (thresholds are the ability levels at which the response to either of 2 adjacent categories is equally likely); (e) threshold differences larger than 1.1 log-odd units and lower than 5.

2) Internal construct validity of the scale was assessed by evaluating the fit of individual items to the latent trait and examining if the pattern of item difficulties was consistent with the model expectation. Fit statistics evaluate the difference between the expected score (i.e., predicted by the model) and the observed score. If the items are contributing useful information about the underlying construct that the scale is sampling, the fit statistics will indicate an acceptable degree of variation.

Fit statistics allow examining the performance of each item on the scale: if the fit statistics on an item indicate less variability in responses than would be expected (“overfitting”), the responses are too predictable (e.g., everyone passes or fails) and the item may not be contributing useful information to the assessment process. Conversely, if the patients’ responses present high variability (“misfitting”), it may be that the item needs to be reconsidered or discarded as it may not be providing information that effectively samples the trait that the scale proposes to assess.

Two types of fit statistics were evaluated: infit and outfit. Infit (inlier-pattern-sensitive fit statistics) is
based on the chi-square statistics with each observation weighted by its statistical information (model variance) and it is more sensitive to unexpected patterns of response by persons on items that are roughly targeted to them, while outfit (outlier-sensitive fit statistics) is based on the conventional chi-square statistics and is more sensitive to unexpected responses to items that are relatively very easy or very hard for the respondent.\textsuperscript{14,20} Acceptable fit for our sample size was defined as MnSq values ranging from .75 to 1.30.\textsuperscript{14} Items with larger values were considered misfitting, while items with smaller values were considered over-fitting.\textsuperscript{19}

3) Reliability was assessed in terms of separation, defined as the ratio of the true spread of the measures to their measurement error.\textsuperscript{14} A person-separation index of 2.0 is considered good, allows the distinction of 3 strata, and corresponds to a reliability of .80, which can be interpreted in a similar way to Cronbach alpha.\textsuperscript{14,21}

4) The principal component analysis of the standardized residuals was performed to investigate the following issues:\textsuperscript{20}
- To have further confirmation of the scale’s unidimensionality, which in this context means that the residuals will be uncorrelated and normally distributed. The following criteria were used to determine whether additional factors were likely to be present in the residuals: (a) a cutoff of 50% of the variance explained by the initial latent trait (Rasch factor) and (b) an eigenvalue of the first contrast <3.
- To analyze the local independence of items. A high correlation (>0.30) of residuals for two items indicates that they may be locally dependent, either because they duplicate some feature of each other or because they both incorporate some other shared dimension.

RESULTS
All questionnaires were completed (no missing responses); four subjects had a minimum score (floor responses).

Factor analysis
The Bartlett Sphericity test was statistically significant (\textit{P} < .001) and the Kaiser-Meyer-Olkin (KMO) test showed a good degree of common variance (KMO = 0.82). Horn PA suggested the presence of one factor with eigenvalues exceeding those from the random data (Table 3). EFA for the 1-factor model showed item loadings for the factor ranging from 0.37 to 0.77 (Table 4). This result was interpreted as a sufficient condition for unidimensionality, and thus for further analysis of the database with Rasch methods.

Rasch analysis
Scale diagnostics demonstrated that the 4-level rating scale complied with the pre-set criteria for category functioning.\textsuperscript{18} Fit statistics revealed that 5 out of the 8 CSD-Ar items fitted the underlying construct that the scale intends to measure. Item 6 "My device is durable" slightly underfitted the module (InfitMnSq = 1.31; OutfitMnSq = 1.39) due to the presence of a few unpredictable responses. In addition, items 1 “My device fits well” and 2 “The weight of my device is manageable” showed overfitting values (with both InfitMnSq and

| Component | Real eigenvalue | Mean random set | 95 percentile random set |
|-----------|-----------------|-----------------|-------------------------|
| 1         | 3.85            | 1.44            | 1.59                    |
| 2         | 1.18            | 1.27            | 1.39                    |
| 3         | 0.80            | 1.14            | 1.23                    |
| 4         | 0.68            | 1.03            | 1.11                    |
| 5         | 0.46            | 0.93            | 1.00                    |
| 6         | 0.41            | 0.83            | 0.91                    |
| 7         | 0.36            | 0.73            | 0.82                    |
| 8         | 0.27            | 0.61            | 0.71                    |

Table 3. Parallel Analysis: Comparison between the size of actual eigenvalues obtained from the principal component analysis with those obtained from a randomly generated data set of the same size and number of variables (500 repetitions).

| Item | Factor loading |
|------|----------------|
| 1. My device fits well | 0.77 |
| 2. The weight of my device is manageable | 0.76 |
| 3. My device is comfortable throughout the day | 0.75 |
| 4. It is easy to put on my device | 0.68 |
| 5. My device looks good | 0.53 |
| 6. My device is durable | 0.37 |
| 7. My device is pain free to wear | 0.52 |
| 8. My skin is free of abrasion and irritation | 0.57 |
OutfitMnSq between .60 and .70).

Figure 1 shows the map of subject-ability and item-difficulty. Subject-ability levels spanned 9.15 logits (from –5.78 to 3.37; average measure –.89), while the item-difficulty estimate spanned 1.54 logits (from –81 to .73). Table 5 shows that a higher item measure (such as that of item 2 “The weight of my device is manageable”) reflects lower scores (using the rating scale from 1 “strongly agree” to 4 “strongly disagree”) and thus higher satisfaction with that orthotics feature (weight). On the contrary, the most difficult item to endorse (higher scores) was item 3 “My device is comfortable throughout the day”.

As for the reliability indices, results were as follows: person-separation index=1.92, person-separation reliability=.79, and Cronbach alpha=.83.

The results of the principal component analysis of the standardized residuals demonstrated that the variance explained by the estimated Rasch measures was fair (54.7%; eigenvalue 9.7), and that explained by the first factor in the residuals was also fair (11.8%; eigenvalue 2.1). Moreover, the correlation between residuals was always lower than .30, except for that between item 1 “My device fits well” and item 3 “My device is comfortable throughout the day” (r=.33).

**DISCUSSION**

The use of psychometrically sound outcome measures is important, as it strongly influences the decisions made by clinicians and researchers to improve the health care service, prescribing, policy making, and expenditure of public funds in the field of orthotics. This study has demonstrated the internal construct validity of the Arabic version of the revised 8-item CSD (CSD-Ar) in patients using various types of orthotics devices in Saudi Arabia. However, the study also confirmed some weaknesses in the CSD scale, mainly in regard to the reliability indicators.

The initial factor analysis suggested the presence of one main factor, which allowed further investigation of psychometric properties to be carried out with RA. RA showed that patients were able to correctly discern among the 4 rating scale options. Fit statistics revealed that item 6 “My device is durable” slightly underfitted the model: the misfitting values belonged to 3 patients with hand injury who—despite a substantial agreement about all the other items—expressed disagreement about the durability of their finger orthoses. In fact, finger orthoses are often made of materials negatively affected by high temperature, water, or mechanical stress; thus a patient could correctly judge the device as lacking durability. Hence, the misfit was not systematic to the item but peculiar to the specific context. Overall, the content validity of item 6 is supported, and thus the item was retained in the questionnaire.

The overfit of item 2 “The weight of my device is manageable” and item 1 “My device fits well” means that the 2 items might be too predictable, and could be redundant for assessing the variable under measurement (i.e., satisfaction with the device) in comparison...
to the other items of the scale. However, overfitting items rarely distort the quality of measurement enough to have any practical metric consequences, and the specific information about manageability and fit of the orthosis seems relevant from a clinical point of view. Anyhow, this finding needs to be confirmed in different populations, devices, and contexts, before considering alternatives to this item choice.

In our study, the hierarchy of item-difficulty showed a similar allocation of items along the scale compared to previous studies. Items regarding weight and fit were most easy to agree with (lower scores), while items addressing comfort were most difficult to agree with. Nevertheless, a detailed comparison with previous studies cannot be made since these studies were based on different populations and devices.

The item-difficulty was reasonably well targeted to the patient-ability in our study sample, as indicated by a mean person measure of -.89. Furthermore, our study showed an acceptable but quite low reliability of the CSD-Ar (person separation reliability=.79; Cronbach alpha=.83). These values favorably compare with those reported in previous studies about CSD. They confirm that the scale is sensitive for distinguishing between just 2 to 3 satisfaction levels (low, medium, and high), which is useful for group decisions but not for everyday clinical application in single patients, where a minimum of .90 is desirable. To enhance the reliability of the scale, further studies should verify the effect of adding items of greater difficulty and/or replacing the present “agree-disagree” type scale with different response options to express agreement or satisfaction.

The principal component analysis of standardized residuals showed a fair unidimensionality (more than 50% of the variance was explained by the Rasch factor) without the presence of a significant second dimension. There was just one correlation between residuals greater than 0.30 (item 1 “My device fits well” with item 3 “My device is comfortable throughout the day”; r=.33): this indicates a negligible local dependency between the two items.

Care should be taken in interpreting our data. First, the CSD-Ar targets prosthetics and orthotics users, but this study was limited to patients using orthotics only. Thus, further studies are needed to confirm the symmetric properties of this outcome measure for patients using prosthetics. In addition, we used a consecutive sampling procedure, and our sample cannot be readily assumed to represent the general population of orthosis users. Although our sample was heterogeneous in terms of age and impairments, the majority of cases were adults with upper limb impairments; thus, patients’ satisfaction was mainly evaluated for upper limb orthotics. Finally, the generalizability of the results is geographically limited, as this study was conducted in only 1 of the 22 Arabic-speaking countries.

In conclusion, this study has confirmed the internal construct validity of the CSD-Ar in patients with various types of orthotics. The CSD-Ar is a promising tool that is applicable to the Arabic culture. The use of such an outcome measure in Arab countries could be of great benefit in improving the services provided by rehabilitation departments. Moreover, our results provide a useful starting point for further analysis and refinement of this outcome measure of PSwO.

Conflict of interest
The authors declare that there is no conflict of interest.

Acknowledgments
The first author thanks the Saudi Arabian Ministry of Higher Education for sponsoring her continuing education. Thanks also to the therapists and those who helped the authors to collect data at King Abdulaziz Medical City Hospital and at a second hospital that wishes to remain anonymous.
REFERENCES

1. Peaco A, Halsne E and Hafner BJ. Assessing satisfaction with orthotic devices and services: a systematic literature review. J Prosthet Orthot. 2011; 23: 95-105.
2. Lindner HYN, Nätterlund BS, Hermansson LMN. Upper limb prosthetic outcome measures: review and content comparison based on international classification of functioning, disability and health. Prosthet Orthot Int. 2010; 34: 109-28.
3. Karmarkar AM, Collins DM, Kelleher A, Cooper RA. Satisfaction related to wheelchair use in older adults in both nursing homes and community dwelling. Disabil Rehabil Assist Technol. 2009; 4: 337-43.
4. Wessels RD, Witte LPD. Reliability and validity of the Dutch version of QUEST 2.0 with users of various types of assistive devices. Disabil Rehabil. 2003; 25: 267-72.
5. Heinemann AW, Bode RK, O’Reilly C. Development and measurement properties of the Orthotics and Prosthetics Users’ Survey (OPUS): A comprehensive set of clinical outcome instruments. Prosthet Orthot Int. 2003; 27: 191-206.
6. Bosmans J, Geertzen J, Dijkstra PU. Consumer satisfaction with the services of prosthetics and orthotics facilities. Prosthet Orthot Int. 2008; 33: 69-77.
7. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000; 25: 3188-91.
8. Jarl GM, Heinemann AW, Hermansson LMN. Validity evidence for a modified version of the Orthotics and Prosthetics Users’ Survey. Disabil Rehabil Assist Technol. 2012; 7: 469-78.
9. Jarl G, Holmefur M, Hermansson LM. Test-retest reliability of the Swedish version of the Orthotics and Prosthetics Users’ Survey. Prosthet Orthot Int. 2014; 38:21-4.
10. Jarl GM, Hermansson LMN. Translation and Linguistic Validation of the Swedish Version of Orthotics and Prosthetics Users’ Survey. Prosthet Orthot Int. 2009; 33: 329-38.
11. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans. 1994, 7: 329.
12. Timmerman ME, Lorenzo-Seva U. Dimensionality assessment of ordered polytomous items with parallel analysis. Psychol Methods. 2011; 16: 209-20.
13. Lorenzo-Seva U, Ferrando PJ. FACTOR: A computer program to fit the exploratory factor analysis model. Behav Res Methods. 2006; 38: 88-91.
14. Bond T, Fox C. Applying the Rasch model: fundamental measurement in human sciences. 2nd ed. Mahwah: Lawrence Elbaum Associates 2007.13.
15. Leung Y-Y, Png M-E, Conaghan P, Tennant A. A systemic literature review on the application of Rasch analysis in musculoskeletal disease — a special interest group report of OMERACT 11. J Rheumatol. 2014; 41: 159-64.
16. Belvedere SL, de Morton NA. Application of Rasch analysis in health care is increasing and is applied for variable reasons in mobility instruments. J Clin Epidemiol. 2010; 63: 1287-97.
17. McAllister S. Introduction to the use of Rasch analysis to assess patient performance. Int J Ther Rehabil. 2009;15: 482-90.
18. Linacre J. Investigating rating scale category utility. J Outcome Meas. 1999; 3: 103-22.
19. Wolfe E, Smith Jr E. Instrument development tools and activities for measure validation using Rasch models: part II—validation activities. J Appl Meas. 2007; 8: 204-34.
20. Linacre J. A user’s guide to Winsteps. Rasch-model computer programs manual 3.75.0. Chicago, IL: http://www.winsteps.com/a/winsteps-manual.pdf, 2012.
21. Fisher WP. Reliability statistics. Rasch Meas Trans. 1992; 6: 238.
22. Linacre J. Redundant items, overtfit and measure bias. Rasch Meas Trans. 2000; 14: 755.
23. Bravini E, Franchignoni F, Ferriero G, et al. Validation of the Italian version of the Client Satisfaction with Device module of the Orthotics and Prosthetics Users’ Survey. Disabil Health J, 2014, in press.
24. Ghoseiri K, Bahramian H. User satisfaction with orthotic and prosthetic devices and services of a single clinic. Disability and Rehabilitation. 2012; 34:1328-32.
25. Bland JM and Altman DG. Statistics notes: Cronbach’s alpha. BMJ. 1997; 314: 572.
#### Appendix 1. Arabic version of the Client Satisfaction with Device module.

| رقم | سؤال | موافق | نظرًا | بذل | لا يوجد
|-----|-------|-------|-------|-----|-------|
| 1.  | تشبيه الأداء التأهيلي/الطرف الصناعي بشكل عام | ✔️ | ✔️ | ✔️ | ✔️ |
| 2.  | وزن الأداء التأهيلي/الطرف الصناعي مقبول | ✔️ | ✔️ | ✔️ | ✔️ |
| 3.  | الأداء التأهيلي/الطرف الصناعي مرير طوال اليوم | ✔️ | ✔️ | ✔️ | ✔️ |
| 4.  | من السهل ارتداء الأداء التأهيلي/الطرف الصناعي | ✔️ | ✔️ | ✔️ | ✔️ |
| 5.  | الأداء التأهيلي/الطرف الصناعي جيد المظهر | ✔️ | ✔️ | ✔️ | ✔️ |
| 6.  | الأداء التأهيلي/الطرف الصناعي جيدة الصنع ومدينه | ✔️ | ✔️ | ✔️ | ✔️ |
| 7.  | لا تسبب الأداء التأهيلي/الطرف الصناعي أي جروح أو نور في البشرة | ✔️ | ✔️ | ✔️ | ✔️ |
| 8.  | لا يسبب ارتداء الأداء التأهيلي/الطرف الصناعي الألم | ✔️ | ✔️ | ✔️ | ✔️ |

**أصبحنا نحناء viewport: بالشك في جودة الأداء التأهيلي/الطرف الصناعي ما مدى صحة العبارات التالية؟ (ضع علامة √ أمام الخيار المناسب)**