**ORIGINAL REPORT**

# ASSESSMENT OF CAPACITY FOR MYOELECTRIC CONTROL: EVALUATION OF CONSTRUCT AND RATING SCALE*

Helen Y. N. Lindner, MSc[1,3], John M. Linacre, PhD[2] and Liselotte M. Norling Hermansson, PhD[3]

*From the [1]School of Health and Medical Sciences, Örebro University, Örebro, Sweden, [2]Department of Occupational Therapy, University of Sydney, Sydney, Australia and [3]Centre for Rehabilitation Research, Örebro University Hospital, Örebro, Sweden*

*Objective:* To examine the construct and rating scale of the Assessment of Capacity for Myoelectric Control, an assessment to evaluate ability in using a prosthetic hand.
*Design:* Cross-sectional study.
*Subjects:* Upper limb prosthesis users with different prosthetic levels/sides and prosthetic experience were included (*n* = 96).
*Methods:* Subjects' assessments with the Assessment of Capacity for Myoelectric Control were collected by 6 raters during their regular hospital visits. Rasch analysis was used, since it allowed an analysis of the data at the item and category levels. Dimension, item hierarchy and item fit statistics were used to examine the construct. Different Rasch parameters were used to examine rating scale structure and its use.
*Results:* The consistency of item difficulties with clinical knowledge and the unidimensionality confirmed that the construct is valid. Two items functioned unexpectedly (misfit), but the misfit was idiosyncratic to the sample, not systematic to the items. The 4-point rating scale usefully differentiated the subjects on the basis of their abilities. The use of category 2 was somewhat redundant.
*Conclusion:* The Assessment of Capacity for Myoelectric Control is a valid assessment that evaluates ability in using a prosthetic hand. Revision of the category 2 definition would improve the functioning of the rating scale.

*Key words:* arm prosthetics, assessment, ability, Rasch analysis.

## INTRODUCTION

In the field of prosthetic rehabilitation, people who are fitted with an upper limb prosthesis are usually offered prosthetic training (1, 2). During training, a new prosthesis user will learn to grasp, hold and release objects with the prosthetic hand. Purposeful tasks are usually given to stimulate the integration of prosthetic use into daily life (3–6). The overall aim of training is to assist the person to achieve maximum functional ability in daily activities. In order to evaluate the person's progress in prosthetic use and to facilitate further goal settings, it is therefore important to assess the user's ability on a regular basis.

The Assessment of Capacity for Myoelectric Control (ACMC) is a standardized clinical assessment designed to assess prosthetic control in myoelectric prosthesis users (4, 7). Since the 2-handed tasks used in the assessments are chosen by the prosthesis users and/or the therapists, ACMC is suitable for upper limb prosthesis users of all ages (4, 8). It can be applied to assess users of different prosthetic levels, i.e. from the shoulder to the hand. The responsiveness of ACMC has been evaluated in prosthesis users over an 18-month period and a change in ability to control the prosthetic hand was detected among these users (7). Inter- and intra-rater reliability has also been evaluated previously in ACMC and the findings indicated that raters with experience in prosthetic rehabilitation made more consistent ratings than those with less experience in this field (9).

The ACMC construct has been evaluated previously using Rasch analysis (7). This analysis orders the assessment items hierarchically according to their difficulty and explores whether the items "fit" the Rasch model. It offers different criteria and parameters to help the researcher to evaluate the validity of an assessment test at the item and category levels (10). One criterion is that the Rasch model expects persons with high ability to obtain higher scores on any item than persons with low ability, and expects any person to obtain lower scores on more difficult items than on easier items (11). Hence, any discrepancy between observed and expected scores is summarized in the item fit statistics and can be used to detect items that are poorly defined or misleading. Another criterion is that the construct of an assessment test is considered valid if all the assessment items measure only one dimension, i.e. data collected should reflect a single underlying construct. Parameters such as separation and reliability indices for persons and items provide information about the range of ability in the sample and the difficulty range in the items. These indices help the researcher to determine whether there are enough items in the assessment test and if there is a wide range of ability in the sample.

Fit statistics for both items and subjects demonstrated an acceptable range in the first study of ACMC (7). The difficulty of all 30 prosthetic items was well-targeted at the average sample

---

ability. In this first study, 64% of subjects were assessed repeatedly (210 assessments from 75 subjects). Since the strengths and weaknesses among these subjects were likely to be repeated several times in the data obtained, the abilities of the prosthesis users in that sample might not give the best picture of the functioning of the items. It was hypothesized that a wider range of ability across the sample might provide a better picture of the functioning of items. Therefore, a further evaluation of ACMC based on an increased number of first-time assessments, i.e. single measures, was considered. Furthermore, factors such as gender and prosthetic side could affect the item difficulties. The evaluation of their effect would increase our knowledge about the items and the ability of prosthesis users.

The rating scale structure, i.e. the number and definitions of the categories, has considerable influence on the quality of the data collected (12). The performance of the 30 ACMC items is rated on a 4-point scale. This rating scale is designed to represent increasing capacities for prosthetic control, that is, each category represents a greater ability than the previous category (13). Based on clinical experience, spontaneity in prosthetic control indicates that the prosthesis user is more confident of his/her ability; hence, the rating scale is designed with a range from "0 = not capable" to "3 = spontaneously capable". The discriminative capacity of a rating scale refers to whether the number of categories in the rating scale is sufficient to differentiate the examined persons on the basis of their abilities (14). The question becomes "Are the 4 ACMC categories sufficient to differentiate the prosthesis users on the basis of their abilities?" Since the ACMC rating scale structure has not been evaluated previously, this needs to be studied. Another concern is whether the raters used the 4 categories in the expected manner. Any misuse of category will affect the quality of the data collected. Rasch analysis allows studies of data at the category level that will help the test developer to determine whether the rating scale structure is appropriate and the use of the rating scale is as expected.

The overall aim of this study was therefore: (*i*) to evaluate the construct of ACMC; and (*ii*) to examine the 4-point rating scale structure and its use. With an increased number of first-time ACMC assessments, specific questions were asked:

- Does this sample provide a wider range of prosthetic ability than was found in the first validity study?
- Do all the items work together to measure a single "prosthetic control" dimension?
- Does the item difficulty hierarchy match clinical knowledge about the difficulty of the items?
- Is this hierarchy influenced by gender and prosthetic side?
- Do all the items function as expected?
- Is the 4-point rating scale appropriately constructed to differentiate between prosthesis users with different abilities?
- Have the 4 rating-scale categories been used in the expected manner?

## METHODS

### Subjects

Ninety-six users of upper limb prostheses participated in this study (55 males, 41 females, congenital deficiency 83, amputation 13, right-sided

39, left-sided 57, age range 2–57 years, mean age 11, median 8 years). Their ACMC assessments were collected between September 2000 and December 2004. These subjects were receiving medical consultation, prosthetic fitting and training at the Limb Deficiency and Arm Prosthetic Centre (LDAPC), Örebro University Hospital, Sweden. All of the subjects were fitted with a myoelectric hand that can be opened and closed voluntarily. Depending on the level of deficiency or amputation, some prosthesis users were fitted with additional body-powered or friction-regulated prosthetic joints (either wrist, elbow or shoulder joints). All subjects were exhibiting a normal developmental pattern, both physically and mentally.

Out of these 96 assessments, 21 were new assessments and 75 had been used in the previous validity study of ACMC. In that study, some subjects were assessed repeatedly, whereas others were assessed only once. Since one assessment per subject was sought in this study, only the "first time" assessment from each subject was retrieved for analysis. The 96 assessments comprised assessments from 22 new prosthesis users (assessments were collected during the first prosthetic fitting) and from 74 subjects who had been wearing a prosthesis for a period of 3 months to 19 years. The project was approved by the local County Council ethics committee review board in Örebro, Sweden.

### Instrumentation

The ACMC consists of 30 functional items grouped into 4 hand use areas: gripping, holding, releasing and co-ordinating. Each item represents a particular hand movement designed to evaluate the ability to use a prosthetic hand. In an ACMC assessment session, the rater identifies the 30 ACMC hand movements by observing how the prosthesis user performs a self-chosen activity. The rater notes down the ratings on the hand movements that he/she can identify on a standard scoring sheet. All items are rated on a rating scale consisted of 4 rating categories. An item is assigned to "category 0 = not capable" when the user cannot perform the item after several attempts. The "category 1 = sometimes capable" rating is given when the unskilful prosthesis user performs an item with the help of verbal or physical guidance. "Category 2 = capable on request" is recorded to an item when a skilful prosthesis user does not carry out the item spontaneously and requires verbal encouragement from the rater for its performance. "Category 3 = spontaneously capable" is noted for an item which the prosthesis user is able to perform skilfully and spontaneously (13).

The ACMC is an assessment based on a persons' performance of tasks that are familiar to the person and accomplished in a natural environment. The person is encouraged to choose any 2-handed task that he/she performs with familiarity with the aim that he/she will feel comfortable while using the prosthetic hand in his/her usual way. For example, a small child may choose to play with different toys and an adult may choose to prepare a simple meal. The reason for using 2-handed tasks is that it is very common for a prosthesis user to perform 1-handed tasks with the sound hand.

### Procedure

In order to ensure consistency of ratings among the raters, the ACMC developer first gave a training course to all the other raters (4 occupational therapists and 2 occupational therapy students). All raters received an administration manual on how to use the assessment (13). The aim of the training course and the administration manual was to help the raters to gain a good understanding of the usage of the assessment. Each ACMC item comes with a definition to instruct the rater how to assess the item. As part of the training, the raters watched video clips about different prosthetic movements and learned how to identify and rate them. During data collection, the students confirmed their ratings with their tutors (the experienced ACMC raters in the same clinic) before the assessments were finalized. This was to ensure the reliability of the students' assessments. Six assessments were collected by the 2 students, and the other 90 assessments were collected by the occupational therapists (OT) working at LDAPC.

The subjects were assessed during their regular visits to the LDAPC for socket change or prosthetic training. The rater identified the ACMC hand movements (30 items) from the subject's performance of any self-

chosen tasks. Examples of the chosen bimanual tasks for assessments were: preparation of a simple meal, making the bed, doing crafts, or playing with different toys. The assessment time was around 30 min. The ratings were then noted on a scoring sheet. The rater would mark an item with "–" on the scoring sheet if this particular item, i.e. the particular hand movement, was not observed during the performance. For example, if the subject had not held any delicate objects during the task performance, then the subject had not performed the item "holding without crushing". This item was then rated with "–", i.e. missing.

*Data analysis*

WINSTEPS® version 3.66.0, Rasch measurement software, was used to analyse the data, using the Rasch rating scale model (15). Rasch analysis is a mathematical technique for calibrating linear logit (log-odds units) measures of item difficulty and person ability from ordinal data. One main purpose of this study was to examine the ACMC construct; therefore stable calibration of item measures was necessary, i.e. the measures obtained from the analysis were stable enough to help us to make inference about the construct, with minimal concern about accidents of the sample. The sample size required to achieve stable item calibrations, i.e. an accuracy of $\pm 0.5$ logits at a 95% confidence interval (CI), ranges from 64 to 144 subjects (16). The sample size in the present study ($n=96$) was thus enough to achieve stable item calibrations.

"Person separation index" was used to determine whether this sample exhibited a wider range of prosthetic ability than was found in the previous validity study of ACMC. The "person separation index" and its equivalent "reliability index" indicate how well ACMC discriminates the ability levels of the persons statistically (10). The separation index was subsequently used to estimate the number of ability strata distinguished by the ACMC items (17).

"Principal components analysis" (PCA) of the residuals was used to examine whether there is a second dimension existing in the unexplained variance after the Rasch dimension is extracted. The comparison between the Rasch factor (the variance explained by the item difficulties) and the first residual factor (unexplained variance by the first contrast) identifies possible multidimensionality.

The person-item map was used to assess the alignment between the subjects and the items. Ideally, the mean person ability and the mean item difficulty should be relatively close to each other (called targeting); and item difficulty range should be able to cover a substantial range of person abilities. Factors that could affect the item difficulty were examined. The 2 chosen factors in this study were gender and prosthetic side. Differential item functioning (DIF) procedure was used to examine whether one subgroup would score higher than the other subgroup on an item. An item has a noticeable DIF if the DIF size $> 0.5$ logits. The calibrated measures differing by 0.5 logits or less have no practical relevance to the measurement (16). The *t*-statistic was used to test the significance difference in the item difficulty measures between subgroups ($p < 0.05$) (15). The impact of DIF on person abilities was examined by comparing the mean person ability of the non-DIF items (excluding the DIF items) with the mean ability of all 30 items.

"Mean-square" (MnSq) and "Z-score standardized" (Zstd) fit statistics were used together to determine whether any item deviated statistically from the expectation of the Rasch model (misfit) (10). The $\chi^2$-based MnSq values indicate the size of the deviation and are formulated to summarize 2 types of unexpected ratings: responses close to an item's difficulty (infit) and responses far away from an item's difficulty (outfit). Outfit MnSq is the mean of the unweighted squared standardized residuals, and infit MnSq is the weighted mean of the information-weighted standardized residuals. The Zstd is the MnSq value standardized to a *t*-distribution with infinite degrees of freedom, i.e. a unit-normal distribution. It is used to estimate the statistical significance of the misfit.

The expected MnSq value for an item is 1.0. Item with MnSq that is lower than 0.5 is interpreted as too little variation in the item response pattern. This perhaps suggests that the item is redundant or measuring areas that are overlapping with other items. This type of item does not threaten the validity of ACMC. Item with MnSq of 1.5 indicates that there is 50% more variation in the observed data than is predicted by the Rasch model.

Item with MnSq higher than 1.5 indicates that the item response pattern has too much variation and considered as misfit. Item with MnSq higher than 2.0 degrades the whole measurement and item removal is recommended (15, 18). The statistical significance of the misfit is expressed as Zstd in WINSTEPS®. The acceptable range is –2.0 to +2.0, i.e. it is within a 2-sided 95% CI for a unit-normal distribution.

The choice of fit statistics and the appropriate numerical range of the fit statistics for determining the "fit" of data to the Rasch model depend on both clinical and statistical factors. To be clinically useful, ACMC has to be a valid assessment for prosthesis users with diverse abilities. Therefore, both infit and outfit item statistics were used to assess how well each item is conformed to the Rasch model. An item with MnSq higher than 1.5 and Zstd higher than +2.0 indicates that the item is misfit and the misfit is statistically significant; perhaps suggesting that the item is poorly defined or misleading. This type of "misfit" is a threat to the validity of ACMC and will be investigated further.

The rating scale structure and its use were examined from several perspectives. Firstly, the "Frequency of Use" of each category indicates how many persons have been rated in that particular category. For stable measurement, at least 10 observations of each category are required (15). Secondly, "Observed Person Measures" should increase from a category representing low ability to one representing high ability (10), e.g. for ACMC the "Observed Person Measure" for category 2 is expected to be higher (indicating more ability) than for category 1 (less ability). Thirdly, "Threshold Measure" should also increase with increasing category number. This indicates that each category in turn is more likely to be observed than any other category as the person ability increases. This is crucial for the effective application of ACMC as a clinically useful diagnostic instrument.

Fourthly, an outfit MnSq for each category was used to examine the consistency of use of the category. The expected MnSq value is 1.0. A rating category with outfit MnSq higher than 1.5 indicates that there is 50% more unexplained variation in the category. A rating category with outfit MnSq higher than 2.0 indicates that highly unexpected ratings were recorded in this category, thus indicating that the category may be contributing data detrimental to the measurement system (10).

## RESULTS

*Range of ability*

A "person separation index" of 5.21 indicated indirectly that there was a wider range of prosthetic ability in this sample than in the first ACMC validity study (person separation index 3.79). The ACMC items have separated the 96 assessments into 7.28 statistically distinct ability levels (strata) on the basis of the subjects' abilities. The equivalent "person reliability index" of 0.97 confirmed that the measures produced by this instrument on this sample are highly reliable (statistically reproducible).

*Dimensionality of Assessment of Capacity for Myoelectric Control*

Statistical decomposition of the variance in data indicates that 30.2% of the variance was explained by the Rasch item difficulties. The first factor in the residuals explained 2.3% of the variance in the data. This first factor was dominated by the contrast between 3 items that measured "holding" and the other items. There is a commonality among these 3 "holding" items, but it is not strong enough to indicate that ACMC is incorporating items measuring 2 different dimensions.

*Relationship between items and subjects' abilities*

All 30-item difficulty measures and all ability measures of the 96 subjects are displayed graphically in the person-item

```
         Logit scale
7   X        +
             |
    X        |
6   XX      T+
             |
    X        |
    X        |
5            +
    X        |
    XX       |
    X        |
4           +T
    XXXX     |
    X        |
    XX      S|   G-repetitive grip, without visual feedback
3   XXX      +   G-adjust force, without visual feedback. R-repetitive release, without visual feedback
    XX       |   G-object towards hand
    XXXXXX|       G-feed hand forward
    XXXX     |   G-without visual feedback
2   XX      +S
    XXXX     |
    XX       |   C-when gripping. R-same time, arms in motion. R-timing, arm is in forward/upward position
    XXXX     |   R-timing, arm is in low position
1   XXXX     +   C-when releasing
    X        |
    XXXX    M|
    XXX      |   R-adjust opening width
0   XX      +M G-repetitive grip. H-in motion, without visual feedback. R-repetitive release
    XX       |   R-without visual feedback
    XXX      |   G-adjust force when gripping. G, R-in any position. H-w/out crushing. H-w/out visual feedback
    X        |
-1  XXX      +
    XXX      |   H-in motion
    XX       |   G-tripod pinch, without support. R-without support
    X        |
-2  X       +S G-whole hand, without support
    XXX     S|   G-tripod pinch, with support
    X        |   H-without support
    XXX      |
-3  XXX      +
    XX       |   H-with support
    XXX      |   G-whole hand, with support
    X        |   R-with support
-4  X       +T
    XX       |
             |
             |
-5  XX       +
           T|
    X        |
-6           +
```
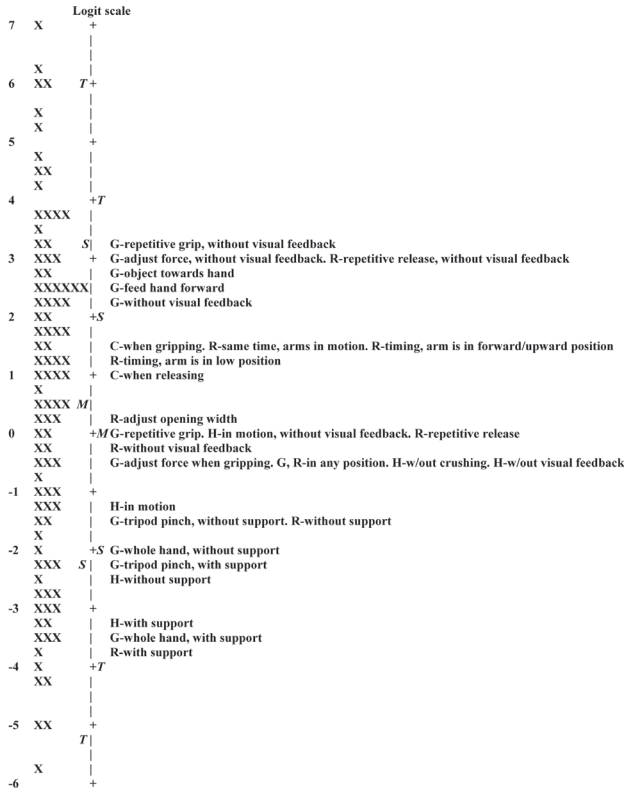
*Fig. 1.* Person-item map: person ability measures in relation to item difficulty measures. X=participant, M=mean for participant ability and item difficulty, S=1 standard deviation (SD) from mean, T=2 SD from mean, G=gripping, R=releasing, H=holding, C=co-ordinating. Person ability measures indicate whether a person is more capable than another, and item difficulty measures indicate whether an item is more difficult than another. All measures are plotted along a shared linear logit measurement scale on which zero is set, by convention, at the average difficulty of the items.

map (Fig. 1). The mean person ability was +0.48 logits with a standard deviation (SD) of 2.81 logits. As seen in Fig. 1, the 2 means are close together, indicating that the difficulty of items targeted well the subjects' abilities. The positive value of the mean subject ability indicates that the subjects, on average, were rated in the upper part of the scale.

The map shows that the item difficulty range is able to cover a substantial distance on the targeted construct, i.e. the ability to control a prosthetic hand. All ACMC items are positioned along the logit scale according to difficulty. As seen in the map, items relating to hand movements performed without visual feedback are the most difficult. Items that need good timing in catching or receiving objects are also relatively difficult. Prosthetic movements that are performed with the arm/hand supported are the easiest. This hierarchy of item difficulty matches the clinical knowledge about the difficulty of different prosthetic movements.

### Factors affecting item difficulty

No item exhibited DIF between subjects with right or left prosthesis. Three items exhibited DIF between males and females. Two items (*repetitive grip without visual feedback, $p=0.02$; repetitive release without visual feedback, $p=0.01$*) were relatively more difficult for males (4.22 and 4.05 logits) than for females (2.48 and 2.05 logits). One item (*adjust force when gripping $p=0.021$*) was more difficult for females (0.26 logits) than for males (–0.75 logits). From the scatter-plot of item difficulties between males and females (Fig. 2), it visualizes clearly that these 3 items fall outside the 95% CI.

The SD of DIF size on gender was 0.35 logits relative to the average item difficulty, which was a small effect when compared with the 7 logit range in the person-item map (Fig. 1). The mean person ability was 0.60 logits, after excluding the 3 DIF items. As compared with the mean person ability for all 30 items (0.48 logits); the person ability differed by 0.12 logits.
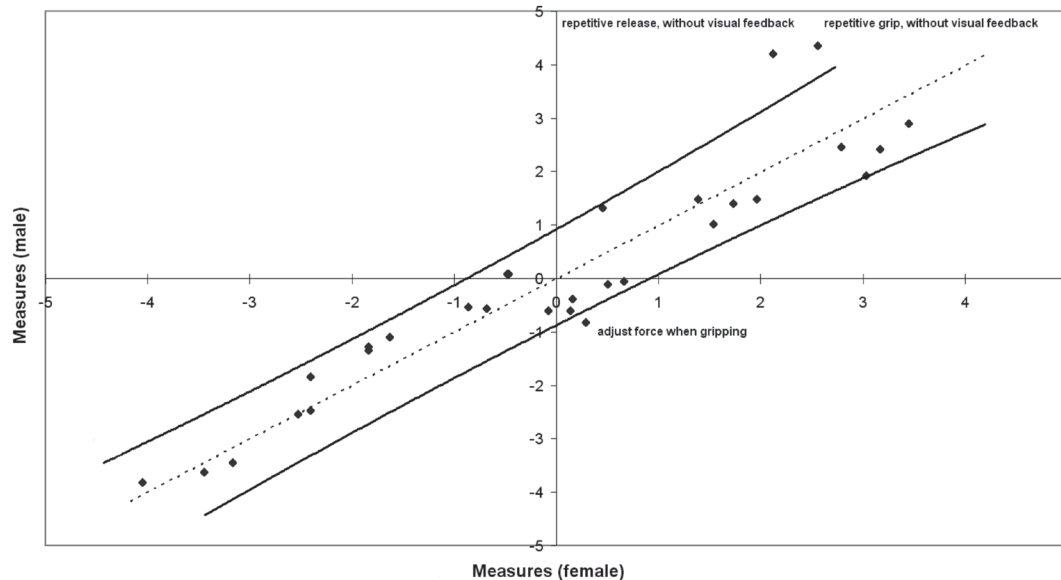


*Fig. 2.* Scatter-plot of item difficulty measures between males and females. Each dot represents an item. The dashed line is the identity line. The solid lines are 95% confidence intervals (95% CI).

Therefore, the effect of DIF on person ability is minimal. When comparing the male and the female item hierarchies, the 2 items that were more difficult for males than for females were both located at the top part of the logit scale, i.e. relatively difficult for both males and females. The item that was more difficult for females than for males was located in the middle range in both the male and female hierarchies.

*Functioning of items*

Each item difficulty measure, its standard error (SE), item fit statistics and the number of missing data for each item are shown in Table I. The SE for each item was acceptably small (mean SE = 0.25) compared with the observed range on the latent variable (see Fig. 1). This indicates that the sample size was large enough to allow stable inferences to be drawn about the items. None of the items had a MnSq value higher than 2.0, indicating that no item needs to be removed from the ACMC strictly on the basis of measurement degradation.

Two items had both infit and outfit MnSq higher than 1.5, but only the infit Zstd were higher than 2.0 (gripping – without

visual feedback, and releasing – same time, arms in motion). In search for an explanation for this, the table of most unexpected responses was examined. This revealed that these 2 items were rated higher than expected in 6 different assessments (3 for each item). Four different raters collected these 6 assessments (not including the OT students) and 3 of them were new assessments, i.e. they had not been analysed in the first study. When these assessments were compared with each other, no obvious association or pattern was detected regarding the reason of limb loss, prosthetic side, age, gender, and tasks.

These unexpected responses were removed in order to see how they would affect the fit for the 2 items. The removal was carried out one response at a time. All the items fitted well (0.5 > MnSq < 1.5) after the removal of the first 3 unexpected responses (2 responses from gripping – without visual feedback and one response from releasing – same time, arms in motion). This implied that these 3 unexpected responses contributed to the "misfit" for both items. This suggests that the misfit in these items was idiosyncratic to this sample, not systematic to the items. It was therefore decided to retain the items in this analysis.

Table I. *Item difficulty measures, the accompanying standard errors and item fit statistics for all 30 Assessment of Capacity for Myoelectric Control (ACMC) items. The items are listed in the order of item difficulty (most difficult to easiest). Misfit items are in bold*

| Item name | Difficulty measure | SE | Infit | | Outfit | | Missing rating per item |
|---|---|---|---|---|---|---|---|
| | | | MnSq | Zstd | MnSq | Zstd | |
| G – repetitive grip, without visual feedback | 3.23 | 0.33 | 1.44 | 1.4 | 1.09 | 0.3 | 64 |
| G – adjust force, without visual feedback | 3.08 | 0.28 | 1.05 | 0.3 | 0.91 | 0.1 | 55 |
| R – repetitive release, without visual feedback | 2.91 | 0.33 | 1.51 | 1.6 | 0.92 | 0.2 | 64 |
| G – object towards hand | 2.69 | 0.26 | 1.18 | 0.8 | 1.58 | 1.0 | 58 |
| G – feed hand forward | 2.54 | 0.25 | 1.00 | 0.1 | 0.92 | 0.1 | 56 |
| **G – without visual feedback** | **2.29** | **0.22** | **1.61** | **2.61** | **1.76** | **1.3** | **36** |
| **R – same time, arms in motion** | **1.59** | **0.26** | **1.74** | **2.6** | **1.57** | **1.1** | **52** |
| C – when gripping | 1.48 | 0.21 | 1.14 | 0.8 | 1.05 | 0.3 | 23 |
| R – timing, arm is in forward/upward position | 1.42 | 0.23 | 1.37 | 1.6 | 1.27 | 0.7 | 53 |
| R – timing, arm is in low position | 1.19 | 0.24 | 1.12 | 0.6 | 0.90 | –0.1 | 52 |
| C – when releasing | 1.09 | 0.22 | 1.15 | 0.7 | 1.30 | 0.8 | 30 |
| R – adjust opening width | 0.22 | 0.19 | 1.07 | 0.4 | 0.76 | –0.6 | 14 |
| H – in motion, without visual feedback | 0.11 | 0.23 | 0.86 | –0.6 | 0.85 | –0.3 | 33 |
| G – repetitive grip | –0.08 | 0.19 | 0.84 | –0.9 | 0.64 | –1.1 | 8 |
| R – repetitive release | –0.08 | 0.19 | 0.89 | –0.6 | 0.67 | –1.0 | 9 |
| R – without visual feedback | –0.21 | 0.26 | 1.42 | 1.7 | 1.19 | 0.5 | 43 |
| H – without crushing | –0.38 | 0.23 | 1.06 | 0.4 | 0.81 | –0.3 | 32 |
| G – adjust force when gripping | –0.38 | 0.20 | 1.28 | 1.4 | 0.90 | –0.1 | 16 |
| H – without visual feedback | –0.40 | 0.24 | 0.83 | –0.8 | 0.93 | 0.0 | 29 |
| G – in any position | –0.54 | 0.20 | 0.80 | –1.1 | 0.86 | –0.2 | 13 |
| R – in any position | –0.60 | 0.20 | 0.88 | –0.6 | 0.89 | –0.1 | 13 |
| H – in motion | –1.23 | 0.20 | 1.01 | 0.1 | 1.05 | 0.3 | 5 |
| G – tripod pinch, without support | –1.42 | 0.20 | 0.86 | –0.7 | 0.50 | –1.1 | 2 |
| R – without support | –1.45 | 0.20 | 1.08 | 0.5 | 0.83 | –0.2 | 1 |
| G – whole hand, without support | –1.95 | 0.20 | 1.15 | 0.8 | 0.89 | 0.0 | 0 |
| **G – tripod pinch, with support** | **–2.34** | **0.22** | **0.46** | **–3.2** | **0.32** | **–1.2** | **13** |
| H – without support | –2.43 | 0.21 | 0.87 | –0.6 | 0.49 | –0.9 | 1 |
| H – with support | –3.20 | 0.23 | 0.80 | –0.9 | 0.37 | –1.0 | 0 |
| G – whole hand, with support | –3.42 | 0.23 | 0.59 | –2.2 | 0.45 | –0.7 | 1 |
| R – with support | –3.75 | 0.24 | 0.62 | –1.9 | 0.24 | –1.2 | 0 |

Infit mean-square (MnSq): weighted mean of the information-weighted standardized residuals. This summarizes the responses close to an item's difficulty.

Infit z-score standardized (Zstd): The infit MnSq value standardized to a t distribution. It estimates the statistical significance of the infit misfit.

Outfit MnSq: unweighted squared standardized residuals. This summarizes the responses further from an item's difficulty.

Outfit Zstd: The oufit MnSq value standardized to a t distribution. It estimates the statistical significance of the outfit misfit.

G: gripping; R: releasing; H: holding; C: co-ordinating; SE: standard error.

Five items had outfit MnSq lower than 0.5, but their outfit Zstd were within the acceptable range –2.0 to +2.0, indicating that the outfit misfit of these 5 items was not statistically significant. Among these 5 items, one item had infit MnSq lower than 0.5 and infit Zstd lower than –2.0 (tripod pinch, with support). These 5 items were the easiest items among all 30 items. They were perhaps too easy for the majority of prosthesis users, and hence there were not many variations in their response strings. These items do not threaten the validity; hence they will not be investigated further.

As listed in Table I, 27 items had "missing ratings". When comparing the missing ratings between subgroups, no obvious association or pattern was detected regarding age or gender.

*Rating scale structure and its use*

Summary statistics for the 4 rating-scale categories are shown in Table II. The "Frequency of Use" of all categories was high, indicating that no category was underused and also that there were sufficient observations of each category for stable inferences to be drawn about the functioning of the rating scale. The "frequency of use" of categories 0, 1 and 2 were fairly even. Category 3 – spontaneously capable was used approximately 3 times more often than any of the other 3 categories (count = 949). This implied that many subjects were spontaneously capable in many items. This is compatible with the previously mentioned result that the mean person ability (0.48) was higher than the mean item difficulty (0).

The results show that the "observed person measures" increased from a low measure for a category representing low ability to a high measure for a category representing high ability. This demonstrates that no collapse of rating categories is needed. The threshold measures increased with the rating category value. This indicates that as the subject's ability is increasing; the raters are most likely to choose 0, then 1, then 2, then 3.

The total range of the 4 category thresholds was 3.13 logits, indicating that the functional range of the rating for any particular item is approximately 4 logits. This was wider than the SD of the person ability measures (2.81 logits), indicating that the 4-point rating scale usefully differentiates the persons on the basis of their abilities.

The outfit MnSq for all categories was ≤1.14, indicating that there was no markedly idiosyncratic use of any of the categories (Table II). The outfit MnSq for "category 2 – capable on request", however, was 0.50, indicating that the use of this category was somewhat redundant. The redundant use of rating category 2 is visualized clearly in Fig. 3. The graph presents the probability associated with the selection of a particular rating category. As the ability increases (moves to the right along the x-axis), the probability of selecting rating category 0 decreases and the probability of selecting category 1 increases, and so on. As compared with the rating category 0, 1 and 3, the probability of selecting rating category 2 was relatively lower (category 2 occupies a relatively small range of ability along the x-axis). Thus, both the outfit MnSq 0.50 and the probability graph suggest that an adjustment of the clinical criterion for "category 2 – capable on request" would perhaps improve the functioning of the whole rating scale.

## DISCUSSION

Whereas the ACMC construct was examined previously with repeated measures, the aim of this study was to examine both the construct and the rating scale with single measures from a larger sample. The results revealed that ACMC is unidimensional and the item difficulty hierarchy is consistent with our theoretical knowledge about the different movements when performing with a prosthetic hand. These 2 important findings again confirmed that ACMC is a valid assessment for measuring ability among users of upper limb prostheses.

The ACMC is an observational tool designed for clinicians to measure changes in prosthetic control in a clinical setting. When testing a clinical instrument such as ACMC, it is always beneficial to use a sample with wide range of ability and to examine how the items in the instrument function with different levels of ability. Although the data set was collected from only one clinic, the ability range was wide enough to test the validity of ACMC. Moreover, in order to avoid local dependence of data due to repeated assessments from the same persons, one assessment per subject was used in this study.
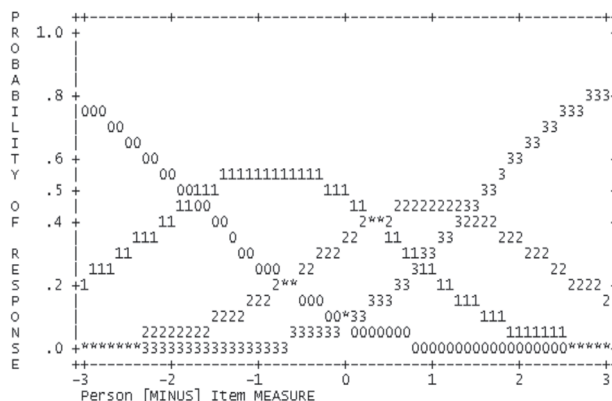
Table II. *Summary statistics for the 4 Assessment of Capacity for Myoelectric Control (ACMC) rating scale categories*

| Category | Frequency of use (%) | Observed person measure | Threshold measure | Outfit MnSq |
|---|---|---|---|---|
| 0 – not capable | 388 (19) | –3.07 | None | 1.14 |
| 1 – sometimes capable | 380 (18) | –0.69 | –1.72 | 1.00 |
| 2 – capable on request | 366 (18) | 1.10 | 0.31 | 0.50 |
| 3 – spontaneously capable | 949 (46) | 3.90 | 1.41 | 1.09 |

Frequency of use: the number of persons rated in that category.
Observed person measure: average person ability measure.
Threshold measure: the difficulty measure between every 2 adjacent categories. It indicates that each category in turn is more likely to be observed than any other category as person ability increases.
Outfit mean-square (MnSq): this is used to examine the consistency of use of the category.



*Fig. 3.* The probability curves of the 4 Assessment of Capacity for Myoelectric Control (ACMC) rating categories. The 0, 1, 2 and 3 category curves on the graph represent the 4 ACMC rating categories.

The PCA result supported the unidimensionality of ACMC, implying that all the items work consistently to measure the control of a prosthetic hand. This important finding confirmed the fulfilment of the ACMC. We designed ACMC with the aim of helping therapists to measure the ability of prosthetic users and set further treatment goals. It is not easy to evaluate the ability change if the assessment involves more than one dimension. Hence, the current psychometric property of ACMC provides an encouraging starting point for measuring change.

From the person-item map (Fig. 1), it is seen that the persons were quite evenly distributed along the linear logit measurement scale. No cluster of persons is observed on the map, implying that the ACMC was sensitive enough even to detect the difference in ability among the 22 new prosthesis users. This sensitivity for differentiation in ability difference can be useful for therapists providing prosthetic training. Any improvement in ability can serve as an indicator of the effectiveness of the prosthetic training provided. Moreover, the person-item map shows the relationship between the range of person abilities and the hierarchy of item difficulties. Hence, this item hierarchy can also be used as a guide for prosthetic training in clinical practice.

Analysis with DIF is always useful to detect item bias. "Repetitive grip, without visual feedback" and "repetitive release without visual feedback" were the 2 items that are more difficult for males than for females. This could be due to the gender difference. In this study, there were more males than females and hence this would further increase the DIF between males and females. It would be interesting to compare the data with other countries in the future and see if these items are still more difficult for males than for females. One might consider that age could affect the item difficulty because the person ability would increase with age. However, on the basis of a previous study (19), the ability development pattern of paediatric prosthetic users and a newly amputated adult (age 39 years) were similar, indicating that older age does not necessarily indicate a higher ability. Other factors such as task types and "delay since prosthetic use" are potential factors that can also affect the item difficulty and further research in the effect of these factors are underway.

The acceptable range for item fit statistics chosen for this study was based on the recommendations from WINSTEPS®, the software we used to analyse our data. However, the use of mean squares or Zstd to identify misfit is debatable. Smith et al. (20) suggests that t-statistics (Zstd in WINSTEPS®) is a more sensitive indicator to identify misfit items than mean squares using a large sample, and Wang (21) even recommended an adjustment of acceptable range for both infit and outfit statistics on the basis of the sample size. A recent study, however, suggested that mean squares statistics are relatively independent of sample size for polytomous data. Furthermore, depending on the type of tests, e.g. observational test or self-rating questionnaires, different fit ranges have been suggested to evaluate item fit (10, 22). Therefore, in the literature it is still uncertain how to determine the most appropriate fit statistic and fit range for evaluating item fit in clinical assessments or other types of tests.

The observed item misfit in this study was due to the fact that some persons received ratings higher than expected. One reason that might have contributed to the item misfit was the influence of task difficulty. A person with low ability might have chosen a very easy and familiar task and received high ratings on difficult items. In the development of another Rasch-derived test, Assessment of Motor and Processing Skills (AMPS) (23), the originators found that persons' measures were dependent on the task performed during the assessments. Thus, it is reasonable to assume that the control of the prosthetic handgrip is easier in some tasks than in others. Hence, it is important that the influence of task difficulty on the functioning of items with standardized tasks should be investigated in future studies. Alternatively, "task difficulty" could be introduced as part of the structure of an extended ACMC, as it is with the AMPS. Further research is needed to confirm the introduction of "task difficulty" into ACMC.

The analysis with a Rasch model is based on the assumption of item independency. The high item reliability and the outfit MnSq values of the easy items raise the issue of item dependency. On the basis of clinical experience, a prosthesis user has to acquire both basic and advanced functions in order to use a prosthetic hand skilfully in different activities. The person has to acquire the basic functions (easy items) before the person can perform the more advanced functions (difficult items). This means that the person who receives a rating on a difficult ACMC item would probably receive a rating on an easy ACMC item, if this easy item is a basic element in that particular difficult item. The 2 items might or might not receive the same rating category, depending on the person's ability. Can we say this is a kind of item dependency? If this is the case, shall we combine those related items to avoid item dependency? If we combine the easy and difficult items that are related to each other, then we would not be able to distinguish prosthesis users who can only perform basic functions from those who can perform both basic and advanced functions. There is a constant tension between the need for accurate clinical information and the requirement of the Rasch model. On the one hand, we hope to develop an assessment that can capture different quality aspects in prosthetic hand function. On the other hand, the assessment has to show good psychometric properties before it can be used clinically. To meet this demand, future development of ACMC could be the collapsing of related items. This could be done without losing the essence of the prosthetic hand movement that we would like to measure.

The rating scales used in several rehabilitation outcome instruments have been examined, and removal or addition of categories has been suggested (24–26). Based on the rating scale analysis in this study, no collapse or addition of rating was found necessary in ACMC. However, the definition of "category 2 – capable on request" needs to be revised, on the grounds of its outfit MnSq and the threshold distances. A person will be rated "category 2 – capable on request" for an item when the person is asked by the rater to perform the item and then performs it skilfully. Thus, the use of the category 2 is very much dependent on both the rater's initiative in giving a verbal request and the prosthesis user's skill. It was very likely

that some raters did not make enough requests during the assessments. One suggestion for the new definition is to omit the rater's request, since this would simplify the use of category 2 and improve the functioning of the rating scale.

In this study the frequency of missing ratings was high (Table I). The reason for missing data was because the person did not use the prosthetic hand as defined in this particular item. As shown in Table I, the number of missing ratings per item tends to increase with item difficulty. This suggests that the prosthesis users could not have demonstrated the items because the items were too difficult for them. If this would be the case, then the prosthetic users would have received the rating "0 – not capable" instead of missing "–". In Rasch analysis, missing data has no effect on the analysis, other than to reduce item precision and reduce reliability. The more data we can collect for an item, the higher the precision and the lower the standard error. However, none of the ACMC items show high standard errors, indicating that all the item difficulties are reliable enough. On the basis of our clinical experience, the item difficulty hierarchy is what we expected, despite the missing values. We have designed ACMC with some easy items, some relatively difficult items and some very difficult items. We do not think that the item difficulties would change dramatically, e.g. an easy item becomes a difficult item if we have all the data, i.e. no missing values. Further research with the design of standardized tasks that contain all ACMC items would not allow missing data. A future comparison of the item difficulties between self-chosen tasks and standardized tasks would provide us with more information on how the tasks and the missing values affect the functioning of ACMC items.

In conclusion, ACMC can be a useful assessment in prosthetic rehabilitation. Revision of one of the rating-scale categories and collapse of several items are suggested for improvement of the instrument. Further research with standard tasks is needed to evaluate the influence of the task difficulty on item functioning. In addition, other reliability tests, such as test-retest reliability, are needed to examine the consistency of a measure over time.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Davids JR, Wagner LV, Meyer LC, Blackhurst DW. Prosthetic management of children with unilateral congenital below-elbow deficiency. J Bone Joint Surg Am 2006; 88: 1294–1300.
2. Egermann M, Thomsen M. Myoelectric prostheses for kindergarten age children. Analysis of first reactions and experiences. Orthopade 2003; 32: 164–169.
3. Hermansson LM. Structured training of children fitted with myoelectric prostheses. Prosthet Orthot Int 1991; 15: 88–92.
4. Hermansson LM, Lindner HYN. Assessment of capacity for myoelectric control – training manual. Örebro: Universitetssjukhuset Örebro; 2008.
5. Hubbard S, Galway HR, Milner M. Myoelectric training methods for the preschool child with congenital below-elbow amputation. A comparison of two training programmes. J Bone Joint Surg Br 1985; 67: 273–277.
6. Lake C. Effects of prosthetic training on upper-extremity prosthesis use. J Prosthet Orthot 1997; 9: 3–9.
7. Hermansson LM, Fisher AG, Bernspang B, Eliasson AC. Assessment of capacity for myoelectric control: a new Rasch-built measure of prosthetic hand control. J Rehabil Med 2005; 37: 166–171.
8. Wright FV. Measurement of functional outcome with individuals who use upper extremity prosthetic devices: current and future directions. J Prosthet Orthot 2006; 18: 46–56.
9. Hermansson LM, Bodin L, Eliasson AC. Intra- and inter-rater reliability of the assessment of capacity for myoelectric control. J Rehabil Med 2006; 38: 118–123.
10. Bond TG, Fox CM, editors. Applying the Rasch model: fundamental measurement in the human sciences. 1 edn. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
11. Rasch G, editor. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1980.
12. Clark HH, Schober MF. Asking questions and influencing answers. In: Tanur JM, editor. Questions about questions: inquiries into the cognitive bases of surveys. New York: Russell Sage; 1992.
13. Hermansson LM. Assessment of capacity for myoelectric control – en metod för att bedöma myoelektrisk kontroll vid utförande av dagliga sysslor [– a method for assessment of myoelectric control in daily life tasks performance]. Örebro: Universitetssjukhuset Örebro; 2000.
14. Linacre JM. Investigating rating scale category utility. J Outcome Meas 1999; 3: 103–122.
15. Linacre JM. Winsteps Rasch Software. 3.66. Chicago: Winsteps; 2008.
16. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994; 7: 328.
17. Wright BD, Masters GN, editors. Rating scale analysis. Chicago: MESA; 1982.
18. Wright B, Linacre J. Reasonable mean-square fit values. Rasch Measurement Transactions 1994; 8: 370.
19. Hermansson LM, Eliasson AC. Clinical application of the Assessment of Capacity for Myoelectric Control. Poster presentation, Hong Kong, ISPO Xl World Congress, 1–6 August 2004. Hong Kong: Hong Kong National Society of the International Society for Prosthetics and Orthotics; 2004
20. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. J Appl Meas 2003; 4: 153–163.
21. Wang WC, Chen CT. Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program for the family of Rasch models. Educ Psychol Meas 2005; 65: 376–404.
22. Linacre JM, Wright BD. Chi-square fit statistics. Rasch Measurement Transactions 1994; 8: 350.
23. Fisher AG, editor. AMPS Assessment of Motor and Process Skills. Fort Collins, CO: Three Star Press; 2003.
24. Denny F, Marshall AH, Stevenson MR, Hart PM, Chakravarthy U. Rasch analysis of the daily living tasks dependent on vision (DLTV). Invest Ophthalmol Vis Sci 2007; 48: 1976–1982.
25. Hsieh CL, Jang Y, Yu TY, Wang WC, Sheu CF, Wang YH. A Rasch analysis of the Frenchay Activities Index in patients with spinal cord injury. Spine 2007; 32: 437–442.
26. Page SJ, Shawaryn MA, Cernich AN, Linacre JM. Scaling of the revised Oswestry low back pain questionnaire. Arch Phys Med Rehabil 2002; 83: 1579–1584.