Assessment of the content validity of Functional Skills Measure after Paralysis with nominal group discussion and revision of its content

Hiroyuki Miyasaka, OTR, MA, Izumi Kondo, MD, PhD, Hiroyuki Kato, OTR, MA, Chikako Takahashi, OTR, Hitomi Uematsu, OTR, Chieko Yasui, OTR, Akina Tani, OTR, Mikiko Miyata, OTR, Noriko Wada, OTR, BA, Toshio Teranishi, RPT, PhD, Yosuke Wada, RPT, MS, Shigeru Sonoda, MD, PhD

1Fujita Memorial Nanakuri Institute, Fujita Health University, Tsu, Mie, Japan
2Fujita Health University Nanakuri Sanatorium, Tsu, Mie, Japan

Correspondence: Hiroyuki Miyasaka, OTR, MA
Fujita Memorial Nanakuri Institute, Fujita Health University, 1865 Hisai-issiki, Tsu, Mie 514–1296, Japan
E-mail: hmiyasak@fujita-hu.ac.jp
Accepted: March 31, 2011
We did not receive any funds for this research work. No benefits in any form have been or will be received from any commercial body related directly or indirectly to the subject of this manuscript.


Original Article

Assessment of the content validity of Functional Skills Measure after Paralysis with nominal group discussion and revision of its content

Hiroyuki Miyasaka, OTR, MA, Izumi Kondo, MD, PhD, Hiroyuki Kato, OTR, MA, Chikako Takahashi, OTR, Hitomi Uematsu, OTR, Chieko Yasui, OTR, Akina Tani, OTR, Mikiko Miyata, OTR, Noriko Wada, OTR, BA, Toshio Teranishi, RPT, PhD, Yosuke Wada, RPT, MS, Shigeru Sonoda, MD, PhD

1Fujita Memorial Nanakuri Institute, Fujita Health University, Tsu, Mie, Japan
2Fujita Health University Nanakuri Sanatorium, Tsu, Mie, Japan

ABSTRACT


Purpose: The purposes of this study were to assess the content validity of Functional Skills Measure after Paralysis (FSMAP) and to revise its content to make it appropriate for use in the clinical setting.

Methods: Eight occupational therapists (OTs) participated in the questionnaire study. Nominal Group Technique (NGT) was used to measure the content validity. Before using the NGT, we set an 80% agreement as the criterion for consent. If the agreement for an item did not reach this level and the item’s content validity was not high enough, we revised its content. We repeated the same assessment for the revised FSMAP by using a questionnaire.

Results: In the first assessment, 8 out of 15 items (including the subitems and descriptions) did not reach the predetermined agreement level. In the second assessment, 1 out of 15 items did not reach the agreement level. We finished the assessment process because we judged that the content validity of FSMAP reached a satisfactory level after the revision of this item.

Discussion: We think that FSMAP should be revised to make it appropriate for use in the clinical setting after assessing its content validity.

Key words: stroke, upper limb hemiparesis, measure, nominal group technique

Introduction

Various evaluative measures have been developed and are currently being used to determine the effect of exercise in patients with upper extremity paralysis after stroke. To evaluate the impairments of paralyzed upper extremities, we used Brunnstrom Stage [1], Fugl-Meyer Assessment [2], Ueda’s 12 level hemiplegic grade [3], Stroke Impairment Assessment Set (SIAS) [4], and Wolf Motor Function Test (WMFT) [5] that took into consideration tasks performed in the daily life. Other studies have used the Motor Activity Log (MAL), which is used to evaluate the frequency of upper extremity paralysis [6].

We developed a Functional Skills Measure after Paralysis (FSMAP), and used it for measuring functional skills that were related to the patient’s approach toward and the use of the upper extremities for daily-life activities, and the nature of this measure, which examined functional skills, was different from that of other above-mentioned measures. Our measures were based on the functional skills of the patients unlike Nagi’s model that was based on the functional limitations of the patients because of their impairments and disabilities [7].

The content of the evaluative measure protocol was affected by the developer’s perspective. The opinion of other specialists was rarely taken into account, and this decreased the validity of the measure in clinical use. Therefore, we attempted to revise FSMAP for its
content validity on the basis of the results obtained from the study by using a nominal group technique (NGT), which is a consensus method for determining functional skills after paralysis.

Methods

1. Participants in NGT
The participants of NGT were 8 occupational therapists (OTs), who had previous experience in working with patients at various stages of recovery in the rehabilitation wards and in the rehabilitation of stroke patients in our hospital. The therapists differed in their experience; while 3 therapists had an experience of 3 years, the remaining 5 had an experience of 1, 5, 6, 8, and 11 years. In addition, a doctor, who was not involved in the development of FSMAP and did not attend the trials, was assigned the role of a facilitator to manage meetings and guarantee fairness for each opinion.

2. FSMAP
In this study, we examined the content validity of the FSMAP, which was originally examined by us as a draft version and comprised 15 items that included tasks performed in the daily life. Each item was subdivided into 3–5 subtests. The items were based on skills that were considered necessary for daily-life activities. The subtests evaluated the functional skills needed to achieve the item, and they were ranked in the presumed order of difficulty. We adopted a rating system in which each test was rated as either possible (score 1) or impossible (score 0) with reference to the description of each test used in the scoring method.

3. Study of content validity
We adopted the NGT to examine the content validity of FSMAP. The consensus method, which is a method used in qualitative study, was used to examine and obtain a consensus for inconclusive issues resulting from insufficient scientific evidence or conflicting arguments. The Delphi method and NGT are examples of the consensus method, and both these methods have been often used in the field of health science.

Initially, we set the agreement level of NGT. For NGT, the median and interquartile range has been commonly used to obtain consensus [9]. Fink et al. [10] introduced various other methods for setting the agreement level and stated that it was necessary to clarify this level before beginning the investigation with a questionnaire. In this study, we set the level at 80% agreement for each question to obtain a consensus that was often used in the Delphi method [11, 12]. In this case, we needed to obtain agreement from at least 6 of the 8 participants. In the questionnaire, although the participants were provided with 4 options, which could later serve as the basis for discussions, we assumed that the consensus would only be reached if 80% of the participants selected the option “Agree.”

5. Procedures of NGT in this study
1) Development of the draft version of FSMAP
To determine the improvement in upper extremity paralysis, we selected 15 items reflecting tasks that involved maneuvering of the upper extremity and divided the items into subtests and provided descriptions for each item on the basis of the collegiate system developed by the research group that included the authors. Details regarding the standardized equipment and the predetermined posture for

Table 1. Feature of consensus method

<table>
<thead>
<tr>
<th>Anonymity</th>
<th>To avoid dominance ; achieved by use of a questionnaire in the Delphi method and private ranking in nominal group technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration</td>
<td>Processes occur in “rounds,” allowing individuals to change their opinions</td>
</tr>
<tr>
<td>Controlled feedback</td>
<td>Showing the distribution of the group’s response (indicating to each individual their own previous response in Delphi)</td>
</tr>
<tr>
<td>Statistical group response</td>
<td>Expressing judgment using summary measures of the full group response, giving more information than just a consensus statement</td>
</tr>
</tbody>
</table>

(Jones J & Hunter D, 1995)

evaluation were included in the manual.

2) Distribution of articles
During the consensus-building process, we distributed articles on the evaluation measures used in Japan and other countries to the participants. This was meant to ensure that the participants had a common concept regarding the content and criterion of the evaluative measure.

3) Design and distribution of questionnaires
We designed the questionnaire focusing on the following 2 points: (1) each item contained subtests that could be used to measure every step from the initiation of the task to its completion and (2) the descriptions of the subtests met the requisite for performing the task. The items in the questionnaire distributed to the participants were rated on a 4-point scale comprising the following options: “Agree,” “Neutral,” “Somewhat disagree,” and “Disagree.” We also prepared a separate column to allow the participants to express their opinions or identify unclear points and suggest improvement plans for the item and the description of FSMAP. We asked the participants to use the original draft of FSMAP and to reflect on the experience of their trial use of FSMAP while answering the questions. The questionnaire was anonymous, and each participant answered the questions in separate rooms.

4) Summary of the questionnaire and its feedback to the participants
We assigned a value to each option in the questionnaire: 1 point for “Agree,” 0.5 points for “Neutral,” -0.5 points for “Somewhat disagree,” and -1 point for “Disagree.” After the participants completed the questionnaire, we determined the individual scores of the participants by adding the scores for each item and gave them feedback on the distribution of the scores and the contents of the column filled by them; then we held the first meeting.

5) First meeting
We presented the results of the questionnaire to all the participants and started proceedings under the supervision of the facilitator. Initially, we gained support from the participants because we adopted the subtests and descriptions without modification and achieved more than 80% agreement. In other words, more than 6 out of 8 participants selected the option “Agree.” Thereafter, we discussed the remaining disagreements with all the participants. We revised the subtests and descriptions on the basis of the discussions.

6) Redistribution of the questionnaire
After the retrial of the modified FSMAP, we redistributed the questionnaire to the participants. The style of answering was similar to that of the first questionnaire. The participants were allowed to change their opinion on the basis of the results of the first questionnaire.

7) Summary of the second questionnaire and feedback to the participants
Like the first summarization, we added the scores for each item in the questionnaire and provided feedback to the participants.

8) Second meeting
The interval between the first and second meeting was around 1 month. As stated below, there was only 1 item that was not agreed upon. Therefore, we cancelled the third meeting after we had obtained agreement after the participants altered their opinions in the second meeting. Figure 1 shows the process of validation.

Results
The results of the first questionnaire suggested that FSMAP did not have sufficient internal validity, because 8 of the 15 items did not fulfill the level of agreement (Table 2).

In the results of the first questionnaire, only 37.5% of participants who were judged agreed on both subtests, “Garments” and “Pants.” This was drastically lower than the predetermined agreement level. With regard to the item, “Garments,” the participants provided 3 comments in the free comment column. The first comment was as follows: “The level of cooperation of the non-paralyzed side was unclear.” The second comment was regarding the description of the subtest: “When the patient fixed the garments, the patient could grasp and pull the (1) front corsage, (2) sleeve, or (3) collar.” The third comment was a question: “Was the difficulty level the same when the patient performed the skill for 1 of the 3 tasks or when the patient performed all of them?” With regard to the item, “Pants,” the following question and description were included in the free comment column respectively: “Isn’t it important for the item to support the body by placing the hand on a platform or bed?” and “The description did not match the order of difficulty.”

In addition to presenting the results of the first questionnaire, we conducted a meeting in which we emphasized on the above-mentioned comments included in the column and modified the description not only for the items, “Garments” and “Pants,” but also for the other items in order to easily comprehend their content. We also improved the descriptions by adding details regarding the settings and criteria used. Figure 2 shows the points that were modified for the item “Garments.”

In the results of the second questionnaire, although the item, “Washing face,” had 75% agreement, all the
other subtests and descriptions had a greater than 80% agreement. The agreement on the subtest of “Garments” was 87.5% and that of “Pants” was 93.8% (Table 2). The comments in the free comment column for “Washing face” stated, “When the patients pooled water with both hands, the degree of overlap of the hands was difficult to measure”. We found that the level of participation of the paralyzed side was unclear; we had not considered this during the first discussion. We discussed the above-mentioned comments and problems during the second meeting and decided to clearly describe the method of use for either the paralyzed or non-paralyzed sides and the criteria to look out for in the subtest and the description.

We judged that the content validity was at a satisfactory level, because the participants agreed to the content of 14 items, and the remaining item could be appropriately modified. Thereafter, we decided to finish the process of validation.

**Discussion**

In this study, we adopted the qualitative research method to examine the content validity of FSMAP. Reports by Palisano et al. [11] and Morris [13] suggested that the content validity could be examined by using the consensus method. In Japan, a few measures have been developed using processes similar to the consensus method. However, during the course of this study, we realized that the advantage of the method for the examination of content validity using the NGT was in the fact that we could repeat the discussions until the content validity of a measure reached the predetermined level.

Quantitative research has been commonly used in the development of a common evaluative measure. Our findings suggested that quantitative and qualitative research were complementary. However, qualitative research was used before quantitative research, because previous studies had used the former research method.

In addition, previous studies have also suggested that qualitative research was useful when validating quantitative research and when pointing out the problems related to societal phenomena from different perspectives [9]. Britten and Fisher [14] suggested that quantitative methods were reliable but not valid, whereas qualitative methods were valid but not reliable. They also stated that generalizations might be made on the basis of the range and diversity of experiences and the formulation of a coherent structure of evidence to explain this diversity, but not on the basis of statistical representativeness. Therefore, the qualitative method has been recently used in a survey conducted by experts to develop a guideline; this suggested the importance of qualitative research in the field of medicine.

In the results of this study, we showed the process of examination of the content validity and modification of the content of FSMAP with regard to the items “Garments” and “Pants” that we presented as the examples of items whose descriptions had to be corrected. Many participants selected the option “Agree” for both subtests and descriptions. Although we created subtests, in which each item was created to measure every step from the initiation of the task to its achievement in the order of difficulty to develop the draft version of FSMAP, various individual differences and a variety of movements in methods and patterns to change clothes were pointed out in the results of the first questionnaire. In other words, the activity of changing clothes was difficult to rank linearly because it involved a variety of movements. Therefore, with regard to the description in “Garments,” we defined the criteria from the beginning of the movement to its completion.

![Figure 1. Process of validation](image)
### Table 2. Results of questionnaire

<table>
<thead>
<tr>
<th>First time</th>
<th>Subtests</th>
<th>Agree</th>
<th>Neutral</th>
<th>Somewhat disagree</th>
<th>Disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of a Cup</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>62.5*</td>
</tr>
<tr>
<td>Manipulation of a PET Bottle</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>56.3*</td>
</tr>
<tr>
<td>Lifting a PET Bottle</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>56.3*</td>
</tr>
<tr>
<td>Holding Paper</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>56.3*</td>
</tr>
<tr>
<td>Tearing Newspapers</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Folding Paper</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>81.3</td>
</tr>
<tr>
<td>Turning over Pages</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>81.3</td>
</tr>
<tr>
<td>Manipulation of a Coin</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>75.0*</td>
</tr>
<tr>
<td>Manipulation of a Bowl</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>50.0*</td>
</tr>
<tr>
<td>Manipulation of a Spoon</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>81.3</td>
</tr>
<tr>
<td>Wearing Garments</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>56.3*</td>
</tr>
<tr>
<td>Manipulation of Buttons</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>81.3</td>
</tr>
<tr>
<td>Wearing Pants</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Washing the Face</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>75.0*</td>
</tr>
<tr>
<td>Washing the Hand</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>56.3*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second time</th>
<th>Subtests</th>
<th>Agree</th>
<th>Neutral</th>
<th>Somewhat disagree</th>
<th>Disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of a Cup</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Manipulation of a PET Bottle</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Lifting a PET Bottle</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>87.5</td>
</tr>
<tr>
<td>Holding Paper</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Tearing Newspapers</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>81.3</td>
</tr>
<tr>
<td>Folding Paper</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>87.5</td>
</tr>
<tr>
<td>Turning over Pages</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Manipulation of a Coin</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>87.5</td>
</tr>
<tr>
<td>Manipulation of a Bowl</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Manipulation of a Spoon</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Wearing Garments</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Manipulation of Buttons</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>87.5</td>
</tr>
<tr>
<td>Wearing Pants</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Washing the Face</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
<tr>
<td>Washing the Hand</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93.8</td>
</tr>
</tbody>
</table>

* Items that did not reach consensus

We set the agreement level at 80% for each question to obtain consensus.
Figure 2. Modification of subtests and descriptions for the item, “Garments”

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Original Draft</th>
<th>Modified Draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garments</td>
<td>(The patient fixes the garment while passing the paralyzed arm through the sleeve (Arm kept in front of the stomach).)</td>
<td>(The patient fixes the garment while passing the paralyzed arm through the sleeve (Arm kept in front of the stomach).)</td>
</tr>
<tr>
<td></td>
<td>(The patient passes the paralyzed hand through the sleeve, while the elbow of the paralyzed side moves beyond the side of the body. The patient passes the hand three times.)</td>
<td>(The patient passes the paralyzed hand through the sleeve, while the elbow of the paralyzed side moves beyond the side of the body. The patient passes the hand three times.)</td>
</tr>
<tr>
<td></td>
<td>(The patient pulls the garments into the pants.)</td>
<td>(The patient pulls the garments into the pants.)</td>
</tr>
<tr>
<td></td>
<td>(The patient fixes the garments by using the paralyzed side while putting off the non-paralyzed side sleeve.)</td>
<td>(The patient fixes the garments by using the paralyzed side while putting off the non-paralyzed side sleeve.)</td>
</tr>
</tbody>
</table>

Description

Starting position/Instrument Layout:

The examiner gives the patient garments with hanger-up sleeves. The starting position, the paralyzed hand is placed on the thigh.

Starting Position:

- The patient fixes the garment while passing the paralyzed arm through the sleeve (Arm kept in front of the stomach).
- While passing the paralyzed arm through the sleeve, if the patient can fix the sleeve of the paralyzed side by keeping it at a position more anterior than the side of the body (the perpendicular line from the ear to the floor), then the score is 1.
- If the patient is unable to fix the garment, the score is 0. If the sleeve moves as if to fix the garment each time, then the score is 1.
- While fixing the garment, the elbow can move beyond the side of the body (10 cm in the perpendicular line from the ear to the floor) three times. If it moves over three times, the score is 0. The performance is deemed to be complete when the wrist comes out of the cuff.
- The patient raises the paralyzed hand to pass it through the sleeve. When the patient passes the paralyzed side arm through the sleeve, if the patient can cooperate with moving or raising that arm, the score is 1. If they are unable to do so, the score is 0. They need not always keep their arm up, and may move their. The performance using an extension pattern may be suitable. The task is deemed to be complete when the entire sleeve has moved out of the cuff.
- While passing the non-paralyzed arm through the sleeve, the patient fixes the garments with the paralyzed hand. When the patient passed their non-paralyzed arm through the sleeve after the paralyzed arm, the patient fixed the garments by grabbing or pinching it with the paralyzed arm. They may hold the garments using the non-paralyzed arm. If the patient can fix the garment until the performance is completed, the score is 1. If they cannot do so, the score is 0.
- The patient raises the garment to pass it through the sleeve. When the patient passes the paralyzed side arm through the sleeve, if the patient can cooperate with moving or raising that paralyzed arm, the score is 1. If they are unable to do so, the score is 0. They need not always keep their arm up, and may move their. The performance using an extension pattern may be suitable. It is deemed to be complete when the wrist has moved out of the cuff.
- While passing the non-paralyzed arm through the sleeve, the patient fixes the garment by grabbing or pinching it with the paralyzed hand.
  - When the patient passed the non-paralyzed arm through the sleeve after the paralyzed arm, the patient fixed the garment by grabbing or pinching it with the paralyzed arm. They may hold the garments using the non-paralyzed arm. If the patient can fix the garment until the performance is completed, the score is 1. If they cannot do so, the score is 0.
  - The patient raises the garment into the pants. The patient moves the paralyzed arm along the body and tucks the garment into the pants. If the patient can fix the garment beyond the upper anterior bias plane, the score is 1. If they are unable to do so, the score is 0. During this performance, the patient may remove the parts using the non-paralyzed side.
  - The patient raises the garments into the pants. While putting off garments with the non-paralyzed arm, the patient can fix the cuff or collar with the paralyzed arm and can arrange them. The patient may hold the garment with the non-paralyzed arm. If the patient is able to do so, the score is 1. If not, the score is 0.

Figures:

- Number of trials: Three. If the patient can complete the performance once within three trials, it is deemed to be successful.
end, and later modified them to minimize the effect of the variations in patterns caused by individual differences. We tried to simplify the visualization of the evaluating movement by using the subtest as a reference, because an analysis of the subtests suggested that the results in which the rate of agreement was much lower than the predetermined level was related to the difficulty level of the movement.

We think that processes such as the group discussion based on the results of the questionnaire, application of the experience of participants working at the clinical setting, and modification of content by consolidating the varying opinions helped in gradually increasing the validity. Van de Ven et al. [16] reported that the NGT and Delphi groups were equally effective and were more effective than the conventional interacting groups when solving a fact-finding problem with no known solutions.

In this study, a high level of agreement from the participants was finally acquired with NGT. The limitation of this study was that the results were acquired through the consolidation of opinions and modification of content by the group engaged in the development of this measure. In the future, we hope to increase the content validity of this measure by using the Delphi method and gathering opinions from external experts in order to enable the use of this measure at other centers.

Acknowledgement

We would like to thank Prof. Eiichi Saitoh and Prof. Yutaka Tomita for providing us with thoughtful suggestions and ample support to complete this work.

References


