Exploring the Use of Factor Analysis to Determine the Relevant Dimensions of Outcome for a Given Population in Rehabilitation Science: A Tutorial

Explorer l'utilisation de l'analyse factorielle pour déterminer les dimensions pertinentes des résultats de la science du rétablissement chez une population donnée : un tutoriel

William E. Hodgetts
Paul Hagler
Sandra Thompson Hodgetts

Abstract

More than ever before, efforts are being made to spend healthcare dollars on interventions that are known to be appropriate and effective for a given patient with a given condition. To determine effectiveness, valid and reliable outcome measures are needed, but selecting which ones to use can be a highly complex task. Factor analysis, though most frequently used for data reduction and scale development, holds great promise as a tool for clinician-scientists to select an optimal combination of outcome measures. This tutorial explores the key theoretical concepts associated with performing a factor analysis with this particular purpose and provides hypothetical examples designed to assist the reader in applying the technique and making sense of the output.

Abrégé

Plus que jamais, des efforts sont déployés pour affecter les fonds du secteur de la santé à des interventions qui seront appropriées et efficaces pour un patient donné souffrant d'un trouble quelconque. Pour déterminer l'efficacité, il faut des mesures de résultats valides et fiables, mais il peut être très complexe de choisir la bonne mesure. L'analyse factorielle, qui sert généralement à la réduction de données et à l'élaboration d'échelles, constitue un outil très prometteur pour les cliniciens-chercheurs qui doivent choisir une combinaison optimale de mesures de résultats. Ce tutoriel explore des concepts théoriques clés liés à la tenue d'une analyse factorielle à cette fin et offre des exemples hypothétiques conçus pour aider le lecteur à mettre en pratique la technique et à comprendre les résultats.

Key Words: : Factor analysis, outcome measures, BAHA, rehabilitation science

William E. Hodgetts, *M.Sc.*
Department of Speech Pathology and Audiology
University of Alberta
Edmonton, AB Canada

COMPRU, Caritas Health Group
Misericordia Community Hospital
Edmonton, AB Canada

Paul Hagler, *PhD*
Department of Speech Pathology and Audiology
University of Alberta
Edmonton, AB Canada

Sandra Thompson Hodgetts, *M.Cl.Sc.,*
Faculty of Rehabilitation Medicine
University of Alberta
Edmonton, AB Canada

Introduction

In health care generally, and in the field of rehabilitation science in particular, a critical question occupies much of what we do. The question is, "How can the benefits of my intervention best be documented?" or, stated differently, "What outcome measures should I use to show that my intervention was successful?" In large part, questions such as these have become increasingly important over the last 20 years due to a shift in our thinking about how to manage health care delivery from models based on the cost of intervention, to models based on the value of intervention (Barbour, 1994; Lanser, 1999).

Cost is still a component of the value-based model; however, it is the cost relative to the benefit (the cost/benefit ratio) that is emphasized. Today's patients ask questions such as, "If I buy the more expensive device, is it really all that much better than the cheaper one?" Today's employers ask, "Can you prove that your intervention has made

a difference in that patient's life?" or "Is there a more effective/efficient intervention?" Finally, third-party payers still want to know, "Is our money being well spent?"

These factors have converged and made it essential for clinicians to demonstrate outcomes of care. As Beck (2000) noted, today's stakeholders are seeking substantive evidence that the intervention and/or rehabilitation provided actually makes a difference in the patient's life. Today, clinical opinion and anecdotal patient reports have ceded their relevance to quantitative outcome measures. Stakeholders will continue to seek proof that the data demonstrates that the appropriate outcome was achieved.

A Potential Tool for Making the Case

Some very real challenges face clinicians who wish to demonstrate the effectiveness of their intervention methods. For example, an audiologist who fits bone-anchored hearing aids (BAHA) might have ten or more variables that measure outcomes such as speech recognition, sound quality, patient satisfaction, and patient quality of life. Similarly, a speech-language pathologist who works with children and adolescents who stutter probably tracks at least ten variables, any one or all of which are generally believed to reflect treatment outcomes. Such variables might include measures of percent syllables stuttered, rate of speaking, and stuttering severity in a variety of contexts in addition to speaker self-perception, satisfaction with treatment, self-efficacy, and perceptions of parents and peers. Which variables are the most salient indices of outcome, and how does one know? Factor analysis affords both clinicians a strategy for demonstrating the outcomes of care, especially when they have a large number of dependent variables of relatively unknown consequence or of questionable importance. Before wading directly into factor analysis, let us establish a common understanding of the phenomenon of multi-dimensional outcomes.

Multi-dimensional Outcomes

The demand for proof of outcomes of care inevitably begs the question, "What constitutes good proof?"[1] Is it acceptable to just pick a valid and reliable outcome measure at will to evaluate a person's post-intervention performance? Should we choose two or three? How do we decide?

First, we should understand the concept of dimensionality of outcome. The argument goes like this. If outcome for a particular health condition were uni-dimensional, then any measurement of outcome in that one dimension would be highly related with all other measurements in that dimension and would allow one to fully understand the outcome for a patient from only one measure. For example, when fitting a person with a hearing aid, one could decide that the outcome measure will be the patient's ability to understand speech in quiet with the hearing aid on. This would assume that the audiologist believes that the patient's understanding of aided speech in quiet is the single most salient dimension of hearing aid outcome and that, even if there are other viable indices of outcomes, they relate so closely to understanding of aided speech in quiet conditions that they would be redundant. In other words, if hearing aid outcome were uni-dimensional, then improved performance on this one measure of speech understanding would allow one to generalize to all other dimensions of outcome. Similarly, if post-treatment fluency were the only salient outcome of stuttering therapy, then a child's response to therapy could be determined simply by calculating the percentage of syllables stuttered. Measures of patients' satisfaction with treatment, their self-efficacy, and the perceptions of their parents and peers would be redundant. Theoretically, if other outcomes such as satisfaction and patient self-perception also are legitimate dimensions of intervention outcome, then one would be able to predict all of them by knowing a single strategically selected outcome variable, such as aided speech understanding in quiet for the BAHA user or post-treatment fluency for the stutterer. Unfortunately, an assumption that one particularly salient dimension can be used to predict all other dimensions is not often valid. For example, it is doubtful that we could take a person's weight and use that to predict height, hair colour and shoe size. All three variables are potentially interesting characteristics of the person, but probably only two of them relate to weight. This is a concept we will return to later.

Indeed, the reality is somewhat more complicated. For any given population in any field of rehabilitation science, there are likely to be several relevant outcome dimensions that need to be measured to accurately characterize the effectiveness of our interventions. To return to our hearing aid example, perhaps the person's speech understanding in quiet is better with the hearing aid, but an ability to hear in quiet conditions and an overall satisfaction with the hearing aid are unrelated or minimally related dimensions for hearing aid users. If so, then failure to measure both will result in an incomplete assessment of intervention effectiveness. It is entirely possible that, despite this person's improvements in speech understanding, there is a more general dissatisfaction with the hearing aid's sound quality (an entirely separate dimension), which may reduce the patient's willingness to use the device, a potentially very expensive and wholly unacceptable outcome. In the case of the child who stutters, it would be indicative of a positive treatment outcome if the percentage of syllables stuttered dropped from 43% pre-treatment to less than 4% post-treatment. However, even with such a dramatic improvement in a widely accepted measure of the effects of treatment, it is possible to also learn that the same child's self-efficacy score remains unchanged or that the parents report little or no improvement in the child's fluency at home.

What is needed is a method of identifying the relevant dimensions of outcome applicable across patient/client populations in rehabilitation science. Factor analysis, although traditionally used by psychologists for data reduction and development of measurement scales, holds great promise for rehabilitation researchers and clinicians as a method that can identify relevant dimensions of intervention outcome (Humes, 1999, 2001, 2003).

Factor analysis

Factor analysis can be considered a family of statistical techniques that try to make sense of complex multivariate data (Everitt & Hay, 1992). In general, factor analysis tries to find an explanation for any relationships found among variables in a large set of measures, and it does this by grouping them into a smaller number of underlying dimensions. Stated differently, factor analysis attempts to determine the degree to which a set of measured variables (outcome measures in our case) cluster together and, in doing so, is said to "extract" the salient dimensions from the larger set of variables. There appears to be considerable laxity in the literature over the use of terms such as 'constructs,' 'factors,' and 'dimensions' (Kline, 1994; Watson, 1998). For the purposes of this paper, we will continue to use the term 'dimensions' to represent phenomena that also are appropriately referred to as underlying 'factors' or 'constructs.'

For illustration, we will use an audiological example. We will adapt the tutorial on factor analysis for hearing-aid outcome measures by Humes (2003) to the special case of Bone Anchored Hearing Aids (BAHA), including modification of the set of outcome measures obtained. In particular, we will make use of the same basic set of three illustrative correlation matrices used by Humes (2003) to explore the range of possible outcomes in the factor analysis of outcome measures. For the BAHA, we might have a hypothetical set of 15 outcome measures for this population. Perhaps we collected 3 measures of speech recognition (SR1, SR2, and SR3), 3 measures of sound quality (SQ1, SQ2, SQ3), 3 measures of hearing aid satisfaction (Satf1, Satf2, Satf3), 3 measures of benefit (Ben1, Ben2, Ben3) and 3 measures of quality of life (QOL1, QOL2, QOL3).

Since we are interested in how the measured variables are related to one another, the first step in implementing a factor analysis is to produce a correlation matrix that looks for a relationship between each variable and every other variable. Table 1 shows a possible correlation matrix for our chosen variables. Tables of this type are

Table 1

Hypothetical correlation matrix for a situation in which there are no clusters of highly related outcome measures.

	SR1	SR2	SR3	SQ1	SQ2	SQ3	Satf1	Satf2	Satf3	Ben1	Ben2	Ben3	QOL1	QOL2	QOL3
SR1	1	0.1	0.0	0.1	0.2	0.1	0.2	0.0	0.1	0.2	0.0	0.1	0.1	0.1	0.2
SR2		1	0.0	0.2	0.1	0.2	0.0	0.1	0.2	0.1	0.2	0.2	0.2	0.1	0.0
SR3			1	0.0	0.2	0.0	0.2	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.2
SQ1				1	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.0
SQ2					1	0.2	0.2	0.0	0.2	0.1	0.2	0.1	0.0	0.1	0.2
SQ3						1	0.0	0.1	0.0	0.2	0.2	0.2	0.1	0.2	0.2
Satf1							1	0.0	0.0	0.2	0.0	0.0	0.1	0.2	0.1
Satf2								1	0.2	0.1	0.1	0.2	0.2	0.2	0.2
Satf3									1	0.2	0.2	0.2	0.0	0.1	0.1
Ben1										1	0.2	0.1	0.2	0.0	0.2
Ben2											1	0.2	0.0	0.2	0.2
Ben3												1	0.2	0.1	0.2
QOL1													1	0.1	0.1
QOL2														1	0.1
QOL3															1

usually generated using data from many subjects.

It can be seen that the correlation coefficients are low for all variables, indicating that none of the outcome measures are closely related to one another. With such weak relationships among these variables, it might be reasonable to conclude that all 15 outcome measures are each measuring something unique. In essence, one could say that each outcome measure in this situation represents its own dimension and, therefore, an accurate characterization of the outcome for this BAHA population would require all of them, an extraordinary case of multi-dimensionality. Obviously, this is a highly unlikely scenario.

A second possible scenario is shown in Table 2. Here the correlation matrix for the 15 outcome measures shows a different extreme insofar as the correlations are high for all variables.

This correlation matrix represents the uni-dimensional situation previously discussed. Since all outcome measures are highly related to one another, knowledge of any one measure would allow us to predict the outcome of any other measure. In other words, if this were the case, we would only need to collect information on one outcome measure to accurately characterize this BAHA population. Since all other outcome measures would be so clearly related to it, they would be redundant, a highly unlikely scenario.

As is usually the case, reality lies somewhere between the extremes. Table 3 depicts a more typical hypothetical correlation matrix derived from our 15 outcome measures.

It can be seen that each measure of speech recognition is highly correlated with the other measures of speech recognition (SR1, SR2 and SR3) but not with any other measures. Similarly individual measures of sound quality (SQ1, SQ2 and SQ3) are highly related to one another, as are the measures of benefit (Ben1, Ben2 and Ben3). However, none of them is highly related to the other outcome measures. Similarly, the measurements of Satisfaction (Satf) correlate with one another, as do the measurements of Quality of Life (QOL). Arguably the most interesting relationships depicted in these data are those observable among Satf and QOL. Such inter-relationships suggest a not-altogether surprising underlying dimension that might be labeled "beneficence" or some other term that the researcher believes will adequately capture the dimension.

The correlations discussed so far have been relatively easy to uncover, since we used only 15 outcome measures for our matrix and intentionally made all of the

Table 2

Hypothetical correlation matrix for a situation in which all variables are highly related to one another.

	SR1	SR2	SR3	SQ1	SQ2	SQ3	Satf1	Satf2	Satf3	Ben1	Ben2	Ben3	QOL1	QOL2	QOL3
SR1	1	0.9	0.8	0.9	0.8	0.7	0.8	0.9	0.8	0.9	0.7	0.9	0.9	0.8	0.9
SR2		1	0.9	0.8	0.9	0.9	0.9	0.7	0.7	0.7	0.9	0.8	0.8	0.8	0.7
SR3			1	0.8	0.8	0.9	0.7	0.8	0.9	0.9	0.8	0.7	0.9	0.9	0.8
SQ1				1	0.7	0.8	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.9
SQ2					1	0.8	0.7	0.7	0.9	0.8	0.9	0.9	0.9	0.8	0.7
SQ3						1	0.9	0.9	0.8	0.9	0.8	0.7	0.8	0.9	0.9
Satf1							1	0.7	0.9	0.7	0.8	0.9	0.7	0.7	0.9
Satf2								1	0.8	0.8	0.8	0.8	0.9	0.8	0.8
Satf3									1	0.8	0.9	0.9	0.8	0.9	0.8
Ben1										1	0.7	0.7	0.9	0.8	0.8
Ben2											1	0.8	0.9	0.7	0.9
Ben3												1	0.7	0.9	0.9
QOL1													1	0.7	0.7
QOL2														1	0.8
QOL3															1

Table 3

Hypothetical correlation matrix for a more typical situation in which there are several distinct clusters of highly related outcome measures.

	SR1	SR2	SR3	SQ1	SQ2	SQ3	Satf1	Satf2	Satf3	Ben1	Ben2	Ben3	QOL1	QOL2	QOL3
SR1	1	**0.9**	**0.8**	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
SR2		1	**0.8**	0.1	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2
SR3			1	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.1
SQ1				1	**0.9**	**0.9**	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
SQ2					1	**0.8**	0.1	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2
SQ3						1	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.1
Satf1							1	**0.9**	**0.9**	0.1	0.1	0.1	**0.9**	**0.9**	**0.8**
Satf2								1	**0.8**	0.1	0.2	0.2	**0.9**	**0.8**	**0.9**
Satf3									1	0.1	0.0	0.1	**0.9**	**0.9**	**0.9**
Ben1										1	**0.9**	**0.9**	0.1	0.1	0.1
Ben2											1	**0.8**	0.1	0.2	0.2
Ben3												1	0.1	0.0	0.1
QOL1													1	**0.9**	**0.9**
QOL2														1	**0.8**
QOL3															1

correlational values starkly different. In fact, the required number of comparisons was small enough that we were able to visually scan the matrix and literally see the patterns. We can factor analyze relatively few outcome measures, 15 in this example, or many outcome measures, 50 to 100, or more if necessary, but virtually never would the relationships be so obvious. Realistically, many more outcome variables would be commonplace, and any relationships would be considerably more subtle for the most part.

As the number of outcome measures increases, the number of comparisons to be made in the matrix increases by the following equation: [n x (n-1) ÷ 2] (Norman & Streiner, 2000). For example, with 15 outcomes measures we needed to scan 105 correlations looking for relationships. If the number of outcome measures increased to 20, the number of correlations we would need to scan would increase to 190. With 25 outcome measures the number would be 300. Obviously, visual scanning of 300 correlations with subtle differences in values for possible underlying dimensions would be nearly impossible. Fortunately, computers can scan any number of correlations with relative ease.

How to Proceed

Think back to our first hypothetical scenario. By scanning the correlation matrix, we could easily tell that none of the variables were related to one another. Intuitively, it made no sense to continue to analyze the correlations for underlying dimensions, because it was obvious that no relationships were evident in the matrix. In a more realistic situation in which we may not be able to see relationships with our eyes, statistical software can be set to work performing similar "checks" to ensure that there are at least some relationships within the matrix before proceeding. Two of the most common tests are Bartlett's test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO).

Hypothetical output from SPSS is shown in Table 4. Without going into the details of how each is calculated, some general rules can be used. The closer the KMO measure is to 1, the better. The KMO is based on squared partial correlations. Therefore, a KMO of 1 would represent perfect correlation while a KMO of 0 (zero) would represent no correlation. Some users look for values greater than .50 for each variable, as well as for the set of variables, but Kaiser (as cited in Norman & Streiner,

2000) suggests a more conservative criterion, that the KMO should be at least .60 if one is to continue to the next step of extracting dimensions. Bartlett's test of sphericity produces a chi-square statistic and tests the null hypothesis that the correlation matrix is made up of diagonal elements equal to 1 and off-diagonal elements equal to 0. The researcher wishes to reject this null hypothesis in hopes of being able to proceed to the next step, extracting dimensions. If the chi square is high and the p-value low (i.e., $p < .05$), it likely is safe to proceed. In other words, low scores from a KMO test or a non-significant Bartlett's F statistic indicate that the variables probably are independent of one another, or at least not dependent enough to be interesting from the point of view of factor analysis.

Table 4

Sample output from SPSS showing two of the most common tests of correlation matrix adequacy

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.888
Bartlett's Test of Sphericity	Chi-Square	6230.901
	df	91
	Sig.	.000

Extracting Dimensions

Assuming that the preliminary "check" suggests it is acceptable to proceed, the next step involves extracting the dimensions from the correlation matrix. Again, the intricacies of the calculations are beyond the scope of this paper.[2] In general, what we are hoping to find is a smaller number of dimensions made up of two or more related variables. Ideally, the dimensions will be unrelated to one another, and will explain much of the variance in the entire set of outcome measures (Gardner, 1995). The principal components analysis (PCA) is the most common type of factor analysis for this purpose. PCA generates a number of potential dimensions and rank orders them in terms of the amount of variance each explains. Table 5 shows sample output from SPSS with 15 outcome measures entered into a PCA. By looking at the cumulative percentage column, it can be seen that the first four dimensions (called "Components" in the table) account for approximately 80% of the variance in the entire solution.

Deciding Which Dimensions to Keep

Despite the computer doing much of the work (e.g., doing a preliminary check, identifying potential dimensions by looking for inter-relationships among variables, and rank ordering dimensions that account for the greatest variance), the researcher must still decide which dimensions to keep. In fact, it could be argued that the number of dimensions to retain in a factor analysis is one of the most important steps in the process (Gorsuch, 1983). Unfortunately, despite its importance, determining which dimensions to retain remains one of the most controversial aspects of factor analysis (Allison, Gorman & Primavera, 1983).[3]

For most statistical packages (i.e., SPSS) the default criterion commonly used is the eigenvalue 1 test. Eigenvalues, which are calculated for each dimension (component), are an index of the variance in the data set, specifically the amount of variance explained by that particular dimension. As mentioned previously, all the outcome measures entered into the factor analysis were converted to standard scores (mean = 0, variance = 1). Revisiting our first scenario in which correlations among the 15 outcome measures were found to be unrelated to one another, it is likely that each outcome measure would account for an equal amount of the variance ($100\% \div 15 = 6.67\%$) in that situation. In terms of eigenvalues, each outcome measure would have an eigenvalue of 1. Conversely, looking back at Table 5, it can be seen that each of the four dimensions that account for a large proportion of the variance in that example, all have eigenvalues greater than 1. According to the eigenvalue 1 test, we should retain these four dimensions, but not the others.

Thus far, this process has been fairly straightforward. What if a solution contained two dimensions with eigenvalues of 1.03 and 0.97 respectively? Is a dimension with an eigenvalue of 1.03 more important than a dimension with an eigenvalue of 0.97? Would you keep the former and discard the latter? That is hard to decide because of the small difference. A useful tool, also available in SPSS, to assist with the decision-making is Cattell's Scree Plot (Norman & Streiner, 2000). There are no equations for this test. In fact, it is simply a plot of the eigenvalues for each dimension. Figure 1 shows the scree plot[4] for the 15-measure example presented in Table 5. Reading from left to right, one will note that, starting with the fifth dimension, the line becomes nearly flat, which means that the fifth and subsequent dimensions account for ever-decreasing amounts of the total variance. Alternatively, the researcher can work backward (from right to left) through the rubble looking for the point at which the line takes a precipitous upward trend. The first dimension above the change from horizontal to vertical is the last of the dimensions that account for the lion's share of the variance. Generally this will coincide with principal components whose eigenvalues are greater than 1, because components with eigenvalues less than 1 account for less variance than the original variable and, therefore, explain relatively little of the total variance.

In this straightforward example, the scree plot helps the researcher see that the first four dimensions would be retained, since there is a steep drop in eigenvalues between dimensions 4 and 5. If dimension 5 approached 1, the researcher might consider keeping 5 dimensions in the

Table 5

Sample output from SPSS demonstrating the percentage of variance explained by each underlying dimension in a PCA solution using 15 hypothetical measures.

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	6.776	39.993	39.993
2	1.998	16.876	56.869
3	1.975	12.227	69.096
4	1.186	10.722	79.818
5	.876	3.889	83.707
6	.672	3.888	87.595
7	.339	2.972	90.567
8	.254	2.14	92.707
9	.199	1.719	94.426
10	.165	1.101	95.527
11	.153	0.992	96.519
12	.130	0.982	97.501
13	.109	0.976	98.477
14	.089	0.811	99.288
15	.078	0.712	100

solution. It is important to note that SPSS uses the eigenvalue 1 as its default test of dimensionality. If, based on analysis of a scree plot, a researcher wishes to retain a dimension with an eigenvalue < 1, the researcher would need to lower the eigenvalue criterion (an option in SPSS) and run the analysis again, so the additional dimension would be included in the analysis.

Determining What the Dimensions Actually Represent

Once a decision is reached regarding which dimensions to keep, the researcher will need to figure out what those dimensions actually represent. Earlier in the third hypothetical BAHA example, we decided that, since all three measures of speech recognition correlated with each other, but not with other measures, they might represent a unique dimension that could be called "speech recognition." Similarly, we might also decide that there is a "sound quality" dimension and a "benefit" dimension. Finally, we might have a dimension called "beneficence", since the satisfaction and quality of life outcome measures correlated well with each other. These hypothetical relationships are depicted in Figure 2.

Again, this illustration was intended to make this task look straightforward. If we look back at our 15-outcome measure example in Table 5, we see that dimension 1 accounts for approximately 40% of the variance, dimension 2 accounts for approximately 16.9% and dimensions 3 and 4 account for only 12.2% and 10.7% respectively. Dimension 1 accounts for disproportionately more variance than do the other dimensions. Statistically, there is nothing wrong with this situation; however, from an interpretation standpoint, disproportionate variance between the dimensions makes it more difficult to decide what a particular dimension actually represents (Gorsuch, 1983). In order to distribute the variance more evenly among the 4 dimensions, factor rotation is used.[5] Another common problem with un-rotated solutions is that they are often hard to interpret because some variables tend to load nearly equally on more than one dimension. This is known as *factorial complexity*. Rotated solutions sometimes reduce factorial complexity and make a factor analysis easier to interpret or more sensible, even if the variance explained by each dimension is similar, and may be necessary to obtain plausible groupings of items within each dimension.

Of the several rotation strategies available in SPSS, the researcher can only determine which one provides the optimal solution by performing all of them. However, most researchers do not perform all possible rotations. Instead, they typically stop with the first rotation that provides a meaningful solution.

Varimax rotation is the most common rotation choice. It is an orthogonal rotation that differentiates the original dimensions by revealing either large or

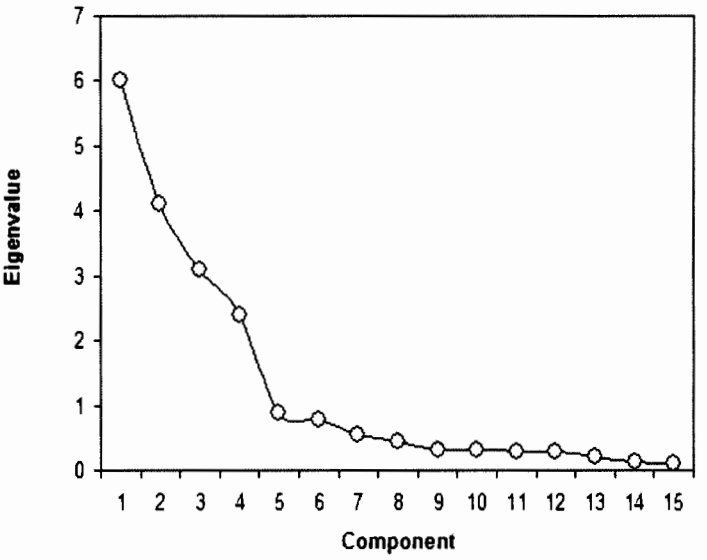

Figure 1. Scree plot of eigenvalues by dimension (component) for a hypothetical 15-measure PCA

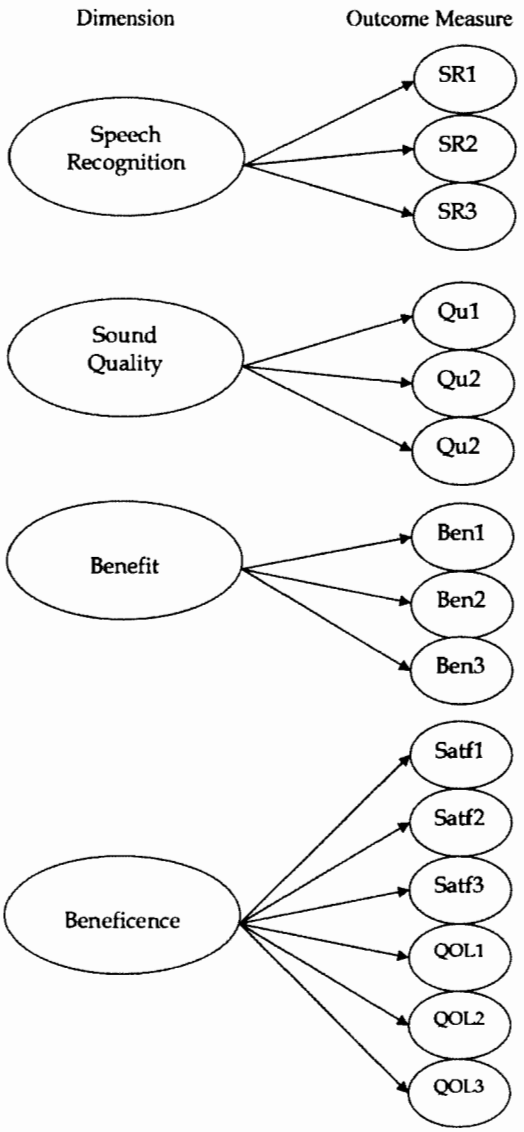

Figure 2. Hypothetical relationships between the 15 outcome measures and the 4 potential underlying dimensions.

small loadings for most variables, making it relatively easy for the researcher to identify all of the variables that belong on a particular dimension. However, when the Varimax rotation does not reveal an optimal combination of dimensions, there are two common alternative strategies that also use orthogonal rotations.

Two common alternative strategies are the Quartimax rotation and the Equimax rotation. The Quartimax rotation tries to reduce the number of variables required to explain each dimension by finding a generic single-dimension solution represented by the most salient variables. Although highly desirable in some research, it generally would not be useful in the current application, which attempts to reveal multidimensional outcomes. The Equimax rotation is a compromise between the Varimax and Quartimax rotations that tries to find a solution with either large or small loadings for most variables. Sometimes this simplifies identification of dimensions and yields a parsimonious set of dimensions. When none of the orthogonal rotations adequately differentiates dimensions, an oblique rotation may be easier to interpret.

Oblique rotations, such as Direct Oblimin and Promax rotations, are non-orthogonal solutions which yield dimensions that, because they are correlated, tend to be harder to interpret. Promax rotations are recommended only for very large data sets, which are not usually available in communication sciences and disorders research. Oblimin and Promax rotations tend to be less useful when one is attempting to simplify and understand an overarching phenomenon having many dimensions, especially when there is an abundance of overlap or correlations among the dimensions.

Once the dimensions have been rotated, the computer will generate a "rotated factor loading matrix." When considering factor loadings, we are asking, "How well does a particular outcome measure load onto each dimension?" For simplicity, the factor loadings can be read and interpreted as though they are correlation coefficients. Technically, they are standardized regression coefficients (Gardner, 1995). The higher the correlation, the better that outcome measure loads onto a particular dimension. Once we have determined which measures load onto a particular dimension, our understanding of the shared nature of those outcome measures begins to take shape. We are now in a position to deduce what the underlying dimension might be and give it a label.

This is illustrated in Table 6. Listed are the rotated factor loadings for each outcome measure on each underlying dimension. By analyzing the factor loadings of each outcome measure, it can be seen that dimension 1 likely represents a "speech recognition" dimension, while dimensions 2 and 3 likely represent "sound quality" and "satisfaction/use" or "beneficence" dimensions respectively.

Indeed, this solution is what we had originally predicted from just the correlation matrix. Now, having gone through the full factor analysis, we have much more definitive evidence that these dimensions actually exist—hypothetically speaking. If this were a real-life example, for our BAHA patients, we could be fairly confident that including at least one measure from each dimension (one measure of speech recognition, one measure of sound quality, one measure of benefit and one measure of either satisfaction or quality of life) might be necessary to accurately assess our intervention. The factor analysis performed its function well; it helped determine the relevant outcome dimensions from a large set of outcome measures for a given population.

Confirmatory factor analysis

Sometimes, when doing factor analysis, one is simply trying to make sense of the data. When that is the case, one is conducting exploratory factor analysis, trying to find out how many dimensions there are and what they look like. In those instances, exploratory factor analysis is really being used as a data reduction technique. Confirmatory factor analysis, on the other hand, requires that we already know, or think we know, what our measures mean, and we want to test our assumptions. Confirmatory factor analysis is used to find out if the number of dimensions and the variables that load on each conform to what is expected on the basis of logic or pre-existing theory. To use it, the researcher must decide a priori what each dimension should be (perhaps even what it should be called) and which variables are the best measures of that dimension. To the extent that the researcher's a priori hypotheses are borne out in the results of the factor analysis, the results are said to be reliable. In confirmatory factor analysis, one tests hypotheses that correspond to prior theoretical notions, which can include the number and nature of factors, but can include much more complex hypotheses, such as the equality of factor pattern matrices across populations or across subsets of the same population represented by more than one sample. In fact, when a sample is of adequate size, we can split it into two approximately equal subsets and perform two factor analyses, one on each set. If the results of the analysis of the second data set confirm the results of the analysis of the first data set, then the confirmatory factor analysis was successful.

Table 6

Rotated factor loading matrix from a hypothetical factor analysis with items loading on each factor circled

	Dimension			
	1	2	3	4
SR1	0.83	0.10	0.11	0.14
SR2	0.88	0.07	0.12	0.15
SR3	0.81	0.06	0.06	0.11
SQ1	0.21	0.87	0.09	0.07
SQ2	0.13	0.92	0.13	0.09
SQ3	0.14	0.89	0.12	0.05
Ben1	0.22	0.11	0.92	0.11
Ben2	0.14	0.12	0.91	0.14
Ben3	0.15	0.06	0.89	0.07
Satf1	0.11	0.09	0.07	0.87
Satf2	0.07	0.13	0.07	0.92
Satf3	0.09	0.12	0.09	0.89
QOL1	0.05	0.07	0.10	0.92
QOL1	0.11	0.07	0.07	0.91
QOL3	0.14	0.09	0.06	0.89

Size of data set

The keen observer will have noticed that no mention has been made of how many outcome measures are needed or how many subjects should be included in a factor analysis. There are no power tables for factor analysis; there are only strongly held beliefs and simple rules of thumb (Gorsuch, 1983; Humes, 2003; Norman & Streiner, 2000). Gorsuch's (1983) text is widely considered to be the most comprehensive treatment of factor analysis. In it, the author suggests that there should be at least four outcome measures for each dimension being assessed. In our hypothetical example, we should have included four outcomes measures for every possible dimension (4 SR, 4 SQ, 4 Satf, 4 Ben and 4 QOL) for a total of 20 outcome measures.

Defining a precise rule for the number of subjects to include in a factor analysis is also difficult (Humes, 2003). Again, Gorsuch (1983) suggests a rule of thumb: The number of subjects should be, at minimum, five times greater than the number of outcome measures in the study and seldom fewer than 100 subjects.

In short, more measures of each dimension are better, and more subjects are always better. It is almost impossible to have too many.

Discussion and Conclusions

Rehabilitation outcome, regardless of discipline and/or health condition, is likely to be multi-dimensional. As such, it is imperative that clinicians attempt to measure the dimensions that are relevant for a given population. There are framework documents and theoretical guidelines that could be used by clinicians to assist them in choosing appropriate outcome measures (e.g. ICIDH-2; WHO, 2003). However, theoretically important factors and empirically important factors for a given population with a given health condition may not be equivalent.

Factor analysis may be considered an alternative or adjunct method of understanding the often complex underlying dimensions of outcomes in healthcare delivery and it is available to clinicians and researchers who desire assistance with the question, "What outcome measures should I use to show that my intervention was successful?"

References

Allison, D. B., Gorman, B. S., & Primavera, L. H. (1983). Some of the most common questions asked of statistical consultants: Our favorite responses and recommended readings. *Journal of Group Psychotherapy, Psychodrama & Sociometry, 46*, US.

Allison, D. B., Gorman, B. S., & Primavera, L. H (1993). Some of the most common questions asked of statistical consultants: Our favorite responses and recommended readings. *Genetic, Social & General Psychology Monographs, 119* (2), 153-185.

Barbour, G. (1994). The role of outcomes data in health care reform. *Annals of Thoracic Surgery, 58*, 1881-1884.

Beck, L. B. (2000). The role of outcomes data in health-care resource allocation. *Ear & Hearing, 21*(4 Suppl), 89S-96S.

Everitt, B. and Hay, D. (1992) Talking about statistics: A psychologist's guide to design and analysis. *Statistical Science, 4*, 394-408

Gardner, R. C. (1995). *Fundamentals of data analysis*. London, ON: University of Western Ontario.

Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.) Hillsdale, NJ: Erlbaum

Humes, L. E. (1999). Dimensions of hearing aid outcome. *Journal of the American Academy of Audiology, 10*, 26-39.

Humes, L. E., Garner, C. B., Wilson, D. L., & Barlow, N. N. (2001). Hearing-aid outcome measured following one month of hearing aid use by the elderly. *Journal of Speech, Language, & Hearing Research, 44*, 469-486.

Humes, L.E. (2003). Modeling and predicting hearing aid outcome. *Trends in Amplification*, 7(2), 42-75.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

Lanser, E. G. (1999). Outcomes and performance measurement: Redefining how healthcare is strategized and delivered. *Healthcare Executive, 14* (4), 20-24.

Norman, G. R., & Streiner, D. L. (2000). *Biostatistics: The bare essentials* (2nd Ed.). Hamilton, ON: B. C. Decker.

Watson, R. (1998). Publishing the results of factor analysis: Interpretation and presentation. *Journal of Advanced Nursing, 28*, 1361-1363.

World Health Organization (2003). Towards a common language for functioning, disability and health. Accessed March 2003 at http://www3.who.int/icf/beginners/bg.pdf.

Footnotes

[1] As it is beyond the scope of this paper, an assumption is made that all outcome measures discussed herein are responsive and have validity and reliability data to support their use.

[2] The interested reader is directed to Gorsuch (1983) for a thorough review of the multivariate general linear model and its relationship to the extraction of dimensions.

[3] See Allison et al. (1983) for a thorough review of this topic.

[4] Scree is the rubble of loose rock that accumulates at the foot of a mountain. In factor analytical terms, it is the rubble of leftover dimensions that accumulate after the important dimensions have been identified.

[5] See Gorsuch (1983) and Norman & Greiner (2000) for additional information on factor rotation.

Author Note

Correspondence concerning this article should be addressed to William E. Hodgetts, Department of Speech Pathology and Audiology, 2-70 Corbett Hall, University of Alberta, Edmonton, Alberta T6G 2G4 Email: bill.hodgetts@ualberta.ca.

Date submitted: March 8, 2005
Date accepted: December 21, 2005