

Focus on Research Methods

Using Measures of Agreement to Develop a Taxonomy of Passivity in Dementia

Cathy L. Antonakos,* Kathleen B. Colling*[†]

University of Michigan School of Nursing, Grants and Research Office,
400 N. Ingalls Street, Room 4243, Ann Arbor, MI 48109-0482
Received 24 September 1998; accepted 8 January 2001

Abstract: A comprehensive taxonomy categorizing passive behaviors in people with dementia was developed and revised through the use of expert raters. The taxonomy was first derived from the synthesis of 15 empirical studies that addressed this phenomenon, then was rated by an expert panel of six nurse-scientists with expertise in dementia and neuroscience research. This article describes the application of two measures of agreement, multiple rater kappa and proportion of agreement for multiple raters, calculated using Stata and SPSS, to evaluate and revise the taxonomy. The method proved useful for estimating the content validity of the taxonomy and provided evidence of stronger agreement among raters for the revised and final forms of the taxonomy. Nurse researchers will find this methodology to be an efficient, practical approach to applying measures of agreement for a variety of purposes, including taxonomy development. © 2001 John Wiley & Sons, Inc. *Res Nurs Health* 24:336–343, 2001

Keywords: expert raters; taxonomy; kappa

Noncognitive behavioral disturbances ranging from passivity to aggression occur frequently over the course of dementing illnesses such as Alzheimer's disease. Prevalence rates for behavioral disturbances are reported to range from 43 to 83% (Ott, Tate, Gorden & Heindel, 1996). Despite recognition that apathy, withdrawal, and underactivity contribute to a significant aspect of behavioral impairment in Alzheimer's disease, relatively little research has been done in this area (Doody, Massman, Mahurin & Law, 1995; Galynker, Roane, Miner, Feinberg, & Watts, 1995). Colling (1999) conducted a descriptive

analysis of 15 empirical studies relating to passive behaviors in dementia. The concept synthesis in that study formed the basis for a taxonomy that classified behaviors associated with passivity in persons with Alzheimer's disease (Colling, 2000). This article reports on the use of six expert raters to estimate the content validity of that taxonomy.

A taxonomy can be defined as a systematically organized framework of labeled sets, groups, or classes linked according to designated criteria (Rasch, 1987). The taxonomy discussed in this article was constructed inductively by listing

This study was supported in part by an NINR Institutional National Research Service Award 5T32NR007074 to Kathleen B. Colling, postdoctoral fellow in neurobehavior at the University of Michigan School of Nursing. We thank David L. Ronis for his valuable assistance in reviewing the manuscript.

Correspondence to Cathy L. Antonakos.

*Assistant Research Scientist.

[†]Lecturer.

behaviors identified as related to passivity in 15 empirical studies of Alzheimer's disease. The behavioral items were examined for similarities and grouped into six inductively derived nonhierarchical categories. These categories included diminution in the areas of cognition, psychomotor activity, feeling emotions, responding to emotions, interacting with persons, and interacting with the environment. The taxonomy was rated by six nurse-scientists with expertise in dementia and neuroscience research, and two measures of agreement, multiple rater kappa and proportion of agreement for multiple raters (Fleiss, 1971), were used to evaluate and refine the taxonomy. This article describes the taxonomy, provides detailed information about the statistics applied, and explains the method used to collect and apply expert ratings.

MEASURES OF AGREEMENT

Measures of agreement are necessary to the conduct of clinical and applied nursing research. These measures estimate the degree and statistical significance of agreement in situations in which two or more judges independently categorize the same set of phenomena (Cohen, 1960). Nurse-scientists find measures of agreement useful in research situations in which the level of measurement is nominal, the categories of the nominal scale can be viewed as exhaustive and mutually exclusive, and reliability in a category coding system is desired. For example, Cohen's kappa (Cohen, 1960) has been used to assess the interrater reliability of a category coding system used to construct a taxonomy of children's coping strategies (Ryan-Wenger, 1992). An extension of the kappa statistic for multiple raters has been used to calculate interrater reliability in the secondary analysis of a qualitative clinical data set (Brennan & Hays, 1992) and to estimate the reliability of patient diagnosis in a clinical setting (Fleiss, 1971). Kukull et al. (1990) used Cohen's kappa to calculate interrater reliability for diagnoses of patients with dementia, comparing scores across three sets of diagnostic criteria. This article presents a method for estimating the content validity of a taxonomy using measures of agreement for multiple raters. The approach provides baseline indicators of interrater agreement regarding the usefulness and clarity of categories and items in the original taxonomy, enabling a comparison with the ratings of a revised version of the taxonomy.

Various statistics are available to evaluate agreement among raters. Percent agreement and phi have been used to estimate interrater reliability, though each has documented limitations (Brennan & Hays, 1992). Percent agreement does not eliminate agreements resulting from chance alone and thus may yield an inflated estimate of agreement. Both phi, a measure of association analogous to a correlation coefficient, and percent agreement can only be used for judgments between two raters (Brennan & Hays, 1992). Cohen's kappa (Cohen, 1960) is also used to measure agreement between two raters. It is appropriate for classifying nominally scaled phenomena, corrects for agreement based on chance alone, and provides an index of reliability that reflects the skill of the raters (Garvin, Kennedy & Cissna, 1988; Topf, 1986). Though Cohen's kappa applies only to the case of two raters, it provides a framework for understanding kappa for multiple raters (Fleiss, 1971).

Cohen's kappa is the ratio of two measures: (a) the difference between the proportion of observed agreement between two raters, I_o , and the proportion of agreement that is entirely by chance, I_e , and (b) the maximum possible difference between observed and chance agreement ($[1 - I_e]$ Fleiss, 1981; Liebetrau, 1983).

$$\kappa = \frac{I_o - I_e}{1 - I_e} \quad (1)$$

A kappa value of 1 indicates perfect agreement beyond chance, whereas a value of 0 indicates that observed agreement is no better than what would be predicted by chance alone.

In studies with more than two raters, Cohen's kappas for all possible pairs of raters are sometimes averaged to obtain an overall kappa for the raters (Kukull et al., 1990). This method may yield a kappa score similar to that of a multiple-rater kappa, but it does not provide a standard error for the kappa. Fleiss (1971) extended Cohen's kappa to measure agreement between more than two raters, presenting an example in which patients' psychiatric illnesses were diagnosed by six independent psychiatrists. Fleiss (1981) suggested a multiple-rater kappa be used for categorical data to measure similarity, concordance, or clustering.

Multiple rater kappa was selected for use in this study because it is a chance-corrected measure of agreement appropriate for evaluating agreement between more than two raters. A category-specific kappa for multiple raters (Fleiss, 1971) is also available for calculating agreement

separately within categories of the taxonomy. However, multiple rater kappa cannot be calculated for an individual phenomenon such as a single descriptor of a passive behavior. To address this limitation of the statistic, we used proportion of agreement for multiple raters (Fleiss, 1971), also known as extent of agreement for multiple raters, to evaluate the usefulness of each item in the taxonomy. Although proportion of agreement scores are not adjusted for chance agreement, higher scores represent stronger agreement among the raters for specific items, with a score of 1.0 indicating perfect agreement among the raters. Later in this article we provide an example that shows how to calculate proportion of agreement for multiple raters using SPSS command syntax, as this statistic is not available in standard statistical software packages.

Landis and Koch (1977) indicated six divisions of multiple rater kappa for interpreting agreement among raters: <0, poor; 0-.2, slight; .21-.4, fair; .41-.6, moderate; .61-.8, substantial; and .81-1.0, almost perfect agreement. Fleiss (1981) used slightly different divisions: <.4, poor agreement; .4-.75, fair to good agreement; and >.75, strong agreement. The formulas for calculating multiple rater kappa are presented in Fleiss (1971). The correct formulas for calculating the standard error for multiple rater kappa are published in Fleiss, Nee, and Landis (1979) and are used in the Stata statistical software package (StataCorp, 1999). For this study the statistical software package Stata version 6.0 was used to calculate the multiple rater kappa and its standard error.

High bias among raters and a high prevalence of a phenomenon have each yielded inflated values for kappa (Banerjee & Fielding, 1997). Though high prevalence was not an issue for this

study, bias among the raters was possible. The kappa values reported in this article are not adjusted for bias, should it exist, as no estimate of bias for these raters was available. A final limitation specific to a multiple-rater kappa is that a standard error is not available if there are unequal numbers of ratings across the set of phenomena, that is, because of missing data (StataCorp, 1999). In this study a standard error could not be computed for the kappa calculated from the initial round of evaluations because of missing data, resulting from some raters not rating all the items in the taxonomy. Standard errors and confidence intervals are reported for kappas calculated for the revised and final versions of the taxonomy. The 95% confidence intervals reported are based on Z scores obtained in the Stata runs.

METHOD

Six PhD-prepared nurse-scientists with clinical and research experience in gerontology independently rated the taxonomy described earlier (Table 1). During the first stage of the analysis, the raters evaluated each behavioral item in the taxonomy, placing it in one of the six inductively derived categories. On the basis of a statistical analysis of the ratings overall (multiple rater kappa), an item-by-item analysis (proportion of agreement among the raters), and other criteria described below, a reduced and revised version of the taxonomy was derived. In the second stage of the analysis, a second round of ratings of the revised taxonomy by the same expert panel was performed to further refine the taxonomy. Although the same panel was used for both reviews of the taxonomy, that the two cycles of

Table 1. Categories of the Original Taxonomy

Number	Label	Definition
1	Diminution of cognition	Lessening of mental processes associated with thinking and knowing
2	Diminution of psychomotor activity	Decrease in spontaneous and purposeful performance of motoric voluntary movements
3	Diminution of feeling emotions	Decrease or absence of the ability to experience sensations associated with an instance of human emotion(s)
4	Diminution in responding to emotions	Decrease in or absence of behaviors that occur following the experience of human emotions
5	Diminution of interactions with persons	Lessening of behaviors indicative of or necessary to acting on or with another or others
6	Diminution of interactions with the environment	Lessening of behaviors indicative of or necessary to acting on or with the physical surroundings

reviews were 6 weeks apart and that substantial revisions occurred after the first round would tend to mitigate the occurrence of rater bias.

The panel of expert raters demonstrated relevant training and qualifications, including a history of published articles in refereed journals, presentations at national meetings, and research experience on the phenomenon of interest (Grant & Kinney, 1992). Although panel members resided in the Midwest, some members had previously lived in other geographical regions of the country, thus increasing the likelihood they could evaluate terms that might have a colloquial connotation (Grant, Kinney, & Guzzetta, 1990). After inviting content experts to serve as panel members, an explanatory cover letter, reviewer instructions, definitions of the categories in the taxonomy, and the passivity category checklist were distributed. Panel members were asked to complete their expert judgments within the week. Because panel members were employed at the same institution, they were asked not to discuss their ratings with any of their colleagues.

Application of Interrater Agreement Measures to Taxonomy Development

Stage one. An instrument presented to six expert raters contained descriptions of the categories of passive behavior in the taxonomy, listed

82 behaviors, and asked the raters to place each behavior in the one category the rater deemed most appropriate. Table 1 provides the descriptions of each of the categories of passive behavior given to the raters for the first round of ratings. In addition, a comment sheet asked raters for information on: (a) their rationale for differentiating between categories, (b) problems encountered with items and categories, and (c) suggestions for additions and/or revisions of categories. This method was consistent with the recommendations of Grant and Davis (1997) for selecting and using content experts for instrument development.

The ratings data were tabulated and manipulated into an appropriate format for analysis, shown in the example of selected behavioral items in Table 2. The data file contains one record per behavioral item, with data in the columns indicating the number of raters who placed each item into a specific category. The kappa for the original taxonomy, calculated using Stata version 6.0, was .43, indicating fair agreement among the raters (Fleiss, 1981). When the number of raters differs across observations, Stata estimates kappa on the basis of all available ratings but suppresses the calculation of test statistics (StataCorp, 1999).

In an effort to refine and improve the taxonomy, individual items were evaluated using proportion of agreement for multiple raters (Fleiss, 1971). The calculations of proportion of agreement for each item in the taxonomy were conducted according to Fleiss (1971) as follows:

Table 2. Data Layout for Calculating Multiple-Rater Kappa and Proportion of Agreement

Selected Behavioral Items	Number of Raters Selecting Category					
	Cognition	Psychomotor Activity	Feeling Emotion	Responding to Emotion	Interactions with Persons	Interactions with Environment
Decreased interest in spending time with friends	0	0	0	0	6	0
Unchanging, "wooden," and "frozen" facial expression	0	2	3	0	0	0
Few expressive gestures when talking	0	0	0	1	0	5
Inability to feel intimacy and closeness	0	1	2	0	0	2
Doing nothing and physically inert	0	5	0	0	0	0
Poor eye contact	0	4	0	0	0	2

N is the total number of items in the taxonomy, n is the number of ratings per item, k is the number of categories in the taxonomy; the subscript i , where $i = 1, \dots, N$, is the item; the subscript j , where $j = 1, \dots, k$, is the category; n_{ij} is the number of raters who placed item i into category j ; and the proportion of agreement among the n raters for the i th item is

$$P_i = \frac{1}{n(n-1)} \left(\sum_{j=1}^k n_{ij}^2 - n \right). \quad (2)$$

The overall proportion of agreement is the average of the P_i 's.

If the number of raters differs across observations, for instance, as a result of missing data, the estimate of the proportion of agreement will be affected. This has implications for decision making. For instance, if an item is rated by four of six raters and three of the four agree on placement of the item, the proportion of agreement score is .50. In this case, because a majority of the six raters did not agree on placement of the item, a rule to retain only items with a majority of raters in agreement could be established. If, in the case of multiple raters, there are missing ratings, it is important for decision-making purposes to consider the number of raters evaluating an item or observation and the proportion-of-agreement score.

Following is the command syntax of SPSS version 9.0 that was used to calculate proportion of agreement for each item in the taxonomy.

*Count for each item the total number of raters who rated that item, by category.

compute nraters = sum(cat1, cat2, cat3, cat4, cat5, cat6).

*Compute the square of the number of raters for each item.

compute sqcat1 = (cat1*cat1).

compute sqcat2 = (cat2*cat2).

compute sqcat3 = (cat3*cat3).

compute sqcat4 = (cat4*cat4).

compute sqcat5 = (cat5*cat5).

compute sqcat6 = (cat6*cat6).

*Compute proportion of agreement for each item (Fleiss, 1971).

compute prop1 = (1/((nraters) * (nraters-1))) * ((sum(sqcat1,sqcat2,sqcat3,sqcat4,sqcat5,sqcat6)) - (nraters)).

For example, because the first item in Table 2 is placed in the same category by all raters, the proportion of agreement score using Equation 1 is:

$$P_1 = \frac{1}{6(5)} 30 = 1.$$

All items with proportion of agreement scores at or above .40, as well as items placed in a specific category by a majority of raters, were retained in the taxonomy. The two emotion categories were combined into a category titled "diminution of emotions" because raters expressed difficulty distinguishing "feeling emotion" and/or "responding to emotion." Redundant or ambiguous wording, as indicated by the raters' comments, was refined and clarified. For example, two items that received low proportion-of-agreement scores (.27)—"Decrease in bathing or showering" and "Decrease in changing clothes daily"—were combined into a single item, "Decrease in performance of activities of daily living." In the second round of ratings the proportion-of-agreement score for this revised item was .67. In another instance an item with a proportion-of-agreement estimate of .28 was retained without rewording as it captured an essential aspect of passive behavior in dementia. That item, "Less willing to try different activities," received a higher estimate in the second round of ratings, .40, and was retained in the final version of the taxonomy. Another item, "Has a 'don't care' attitude," was dropped despite a higher proportion of agreement score (.60) because it could have multiple meanings or interpretations depending on the emotional motivation for the behavior. Thus a combination of comments from the expert raters, judgments by the principal investigator, and statistical information was used to decide whether to revise or retain items, though statistical information held sway for the vast majority of items dropped from the taxonomy. The reduced and revised taxonomy contained 63 items, all of which appeared more likely to fall clearly into one of the category groups.

Stage two. The same six experts on the panel conducted a second round of independent ratings of the revised taxonomy 6 weeks after the first round. Multiple rater kappa and proportion of agreement scores were calculated using the methods described for Round 1. The kappa for this version of the taxonomy was .61, indicating fair to good (Fleiss, 1981) or substantial agreement (Landis & Koch, 1977) beyond chance. The 95% confidence interval for the kappa is .58-.65.

An item-by-item analysis using proportion of agreement scores led to the deletion of nine additional items from the taxonomy, including

Table 3. Number and Percentage of Items on Which Raters Agree

Raters in Agreement ¹	Number and Percentage of Items by Category ²									
	Cognition		Psychomotor Activity		Emotions		Interactions with Persons		Interactions Environment	
	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)
4	4	50.0	3	37.5	1	5.6	3	23.1	3	42.9
5	1	12.5	3	37.5	6	33.3	2	15.4	2	28.6
6	3	37.5	2	25.0	11	61.1	8	61.5	2	28.6
Total	8	100.0	8	100.0	18	100.0	13	100.0	7	100.0

¹Number of raters in agreement out of a maximum of six. None of the items had less than four raters in agreement.

²The categories contain unequal numbers of items.

one item inadvertently printed twice in the revised version of the taxonomy. Intrarater agreement for all six raters was 100% across the two occurrences of this item, indicating consistency of the judges in their own decision making. In general, items were retained in this 54-item version of the taxonomy if four or more raters agreed on their placement, corresponding to extent-of-agreement scores of .40 or above. To summarize, Table 3 indicates items on which the raters agreed by number and percentage. All six raters agreed on placement of 48% of the items in the taxonomy. Five raters agreed on placement of 26% of the items, and four raters agreed on placement of the remaining 26% of the items. The overall kappa for the 54-item version of the taxonomy was .70. This kappa also demonstrated statistically significant agreement among the raters ($p < .001$). The 95% confidence interval for the kappa was .66–.73.

Additional statistics were calculated to provide more information about the taxonomy and the rating process. The Cohen's kappa was calculated for all possible pairs of the six raters (a total of 15 kappas) for the final version of the taxonomy using SPSS version 9.0 (Table 4). These kappas ranged from .60 to .83, reflecting moderate to strong agreement between all pairs of raters (Fleiss, 1981; Landis & Koch, 1977). The strength of agreement among raters within each

category of the taxonomy (Fleiss, 1971) also was calculated for the final version of the taxonomy using Stata version 6.0 (Table 5). These kappas range from .55, for diminution of psychomotor activity, to .82, for diminution of emotions. The lowest kappas indicate categories with need for further refinement.

DISCUSSION

The method applied proved useful for estimating the content validity of the taxonomy and provided evidence of stronger agreement among the raters for the revised forms. The method used suggestions by Goldsmith (1981) and Holsti (1969) to define the task as a series of dichotomous decisions, provide clear definitions of the categories, combine categories as needed, and use several raters. The two-stage method enabled validation of the experts' initial work to improve the taxonomy. In the first phase it was important to include a mechanism for the raters to provide feedback on their impressions in the form of comments that could be used to refine the instrument for the second phase of evaluation. During the second phase redundancy in the categories was reduced, wording of specific items was refined and clarified, and items with low proportion-of-agreement scores were eliminated, yielding items that tended to fall more clearly into one of the categories of the taxonomy. Throughout the total process it was helpful for the researcher to remain objective and impartial during the evaluation of the taxonomy by allowing each expert the freedom to place items into the category of choice.

The use of baseline statistics, indicating the strength of the taxonomy, guided the researcher in the development of an improved taxonomy and yielded empirical validation that indeed the

Table 4. Cohen's Kappas for All Pairs of Raters

	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6
Rater 1	.64	.78	.78	.81	.79
Rater 2	—	.64	.64	.60	.60
Rater 3		—	.67	.83	.67
Rater 4			—	.69	.62
Rater 5				—	.74

Table 5. Description of Final Taxonomy

Category	Category-Specific Kappa (95% CI)	Representative Items
Cognition (8 items)	.67 (.60-.73)	Decreased ability to be goal directed Decrease in ability to imagine. Demonstrates less intellectual curiosity
Psychomotor Activity (8 items)	.55 (.48-.62)	Slowness of movements Decrease in gross motor movements Decrease in performance of activities of daily living.
Emotions (18 items)	.82 (.75-.89)	Inability to feel intimacy and closeness Less enthusiastic Diminished emotional responsiveness Less cheerful than usual.
Interactions with Persons (13 items)	.74 (.67-.80)	Decreased interest in spending time with friends Poor eye contact with others Displays disinterest to others Less responsive when interacting with others
Interactions with Environment (7 items)	.61 (.54-.68)	Seems unmoved by environment Avoids a stimulating environment Seems to be withdrawing from activities

taxonomy did become better through revision. The second stage of ratings by experts is evidence of this improvement. Through continued development the taxonomy has become more likely to meet Reynold's (1971) criteria: exhaustive categories, mutually exclusive and consistent with concepts that describe, explain, predict, and lead to an understanding of the phenomenon. The categories and items of the taxonomy have been used to develop an instrument called the Passivity in Dementia Scale (PDS), intended to measure passive behaviors in people with dementia. Estimates of the validity and reliability of the PDS, along with substantive findings of the research now being conducted on caregivers' and patients' assessments of passive behavior in early dementia, are planned for a future publication. There are many references about methods for estimating reliability and validity, such as Lynn's (1986) content validity index, but fewer sources exist regarding the often less structured initial work in scale development. This article has provided a comprehensive methodology for structuring the early work related to taxonomy development.

REFERENCES

Banerjee, M., & Fielding, J. (1997). Interpreting kappa values for two-observer nursing diagnosis data. *Research in Nursing & Health*, 20, 465-470.

Brennan, P.F., & Hays, B.J. (1992). The kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. *Research in Nursing & Health*, 15, 153-158.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Colling, K.B. (2000). A taxonomy of passive behaviors in people with Alzheimer's disease. *Journal of Nursing Scholarship*, 32, 239-244.

Colling, K.B. (1999). Passive behaviors in Alzheimer's disease: A descriptive analysis. *American Journal of Alzheimer's Disease*, 44(1), 27-39.

Doody, R.S., Massman, P., Mahurin, R., & Law, S. (1995). Positive and negative neuropsychiatric features in Alzheimer's disease. *Journal of Neuropsychiatry*, 7(1), 54-60.

Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378-382.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Fleiss, J.L., Nee, J.C.M., & Landis, J.R. (1979). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72, 323-327.

Galynker, I.I., Roane, D.M., Miner, C.R., Feinberg, T.E., & Watts, P. (1995). Negative symptoms in patients with Alzheimer's disease. *American Journal of Geriatric Psychiatry*, 3(1), 52-59.

Garvin, B.J., Kennedy, C.W., & Cissna, K. (1988). Reliability in category coding systems. *Nursing Research*, 37, 52-55.

Goldsmith, J.W. (1981). Methodological considerations in using videotape to establish rater reliability. *Nursing Research*, 30, 124-127.

- Grant, J.S., & Davis, L.L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, 20, 269–274.
- Grant, J., & Kinney, M. (1992). Using the Delphi technique to examine the content validity of nursing diagnoses. *Nursing Diagnosis*, 3, 12–22.
- Grant, J., Kinney, M., & Guzzetta, C. (1990). A methodology for validating nursing diagnoses. *ANS: Advances in Nursing Science*, 12(3), 65–74.
- Holsti, O.R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.
- Kukull, W.A., Larson, E.B., Reifler, B.V., Lampe, T.H., Yerby, M., & Hughes, J. (1990). Interrater reliability of Alzheimer's disease diagnosis. *Neurology*, 40, 257–260.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Liebetrau, A.M. (1993). *Measures of association*. Newbury: Sage.
- Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382–385.
- Ott, B.R., Tate, C., Gordon, N., & Heindel, W. (1996). Gender differences in the behavioral manifestations of Alzheimer's disease. *Journal of the American Geriatrics Society*, 44, 583–587.
- Rasch, R.F.R. (1987). The nature of taxonomy. *Image: The Journal of Nursing Scholarship*, 19, 147–149.
- Reynolds, P.D. (1971). *A primer in theory construction*. Indianapolis: Bobbs-Merrill.
- Ryan-Wenger, N. (1992). A taxonomy of children's coping strategies. *American Journal of Orthopsychiatry*, 62, 256–262.
- StataCorp. (1990). *Stata Statistical Software: Release 6.0*. College Station, TX: Stata Corporation.
- Topf, M. (1986). Three estimates of interrater reliability for nominal data. *Nursing Research*, 35, 253–255.