Psychometric characteristics of the Everyday Problems Test (EPT)

Sherry L. Willis

Department of Human Development and Family Studies

Pennsylvania State University

Michael Marsiske

Max Planck Institute for Human Development and Education

Running head: PSYCHOMETRIC CHARACTERISTICS OF THE EPT

Abstract

The Everyday Problems Test (EPT) is a measure designed to assess the cognitive competence of older adults in tasks of daily living. The present study administered the EPT to 421 subjects representing a broad age range (54-95 years) and educational distribution (4-22 years). Participants were re-contacted after one year, and 291 subjects returned for retesting. Analysis of EPT data revealed that the EPT could be described by seven highly related domains, which roughly corresponded to seven Instrumental Activities of Daily Living (IADL) categories. Internal consistency and one-year test-retest stability of the EPT was high, and the measure discriminated between young-old and old-old subject groups. Thus, it is argued that the EPT provides a reliable way of assessing the everyday cognitive competence of older adults.

Psychometric characteristics of the Everyday Problems Test (EPT)

Although there has been a growing research interest in the cognitive performance of older adults in real-world contexts (e.g., Hess, 1990; Poon, Rubin & Wilson, 1989; Rogoff & ave, 1984; Sinnott, 1989; Sinnott & Cavanaugh, 1991; Sternberg & Wagner, 1986) relatively little progress has been made in defining what the critical task properties of everyday cognition are. The lack of theoretical consensus on what constitutes a valid set of criterion tasks for the measurement of everyday cognition is reflected in the many different empirical approaches that have been taken (e.g, Sinnott and Cook, 1989).

Willis and Schaie (1986, in press) and Willis (1991) have argued for a critical task definition of everyday cognition. This perspective argues that the important categories of everyday functioning that must be studied are those which predict the maintenance of viability and independence into old age. In their most recent empirical work, Willis and associates have suggested that the Instrumental Activities of Daily Living (IADL) framework of Lawton and Brody (1969) may provide a useful heuristic for the identification of this core set of critical tasks. The IADLs are widely accepted as crucial tasks needed for the maintenance of independent functioning in the community (Allen & Allen, 1987; Branch & Jette, 1982; Fillenbaum, 1985; Lawton, 1987; Wiener & Hanley, 1990; Wolinsky et al., 1983). Impaired IADL competence has been found to be associated with higher levels of mortality (Fillenbaum, 1985; Koyano et al., 1989; Manton, 1988).

Although IADL competence may be an important component of older adults'

maintained viability and adaptive functioning in the community, there have been problems with its measurement. Many studies reporting IADL assessment have used self-report measures with restricted variability (Fillenbaum, 1985). Furthermore, self-report assessments of IADL competence tend to overestimate actual functional levels (Kuriansky, Gurland, & Fleiss, 1976; Little, Hemsley, Volans, & Bergman, 1986; Loewenstein et al., 1989), items tend to be interpreted differently by different subjects (Jobe & Mingay, 1990), and there is only modest agreement between self- and proxy ratings of functional competence (Kuriansky et al., 1976; Marsiske, 1992; Rogers & Holm, 1990).

Attempting to integrate the research traditions on everyday cognition and on the functional competence of older adults, the present study reports on the development and characteristics of a performance-based measure of everyday cognitive competence with instrumental activities of daily living. The Everyday Problems Test (EPT) (Willis, 1990) was designed specifically for use with older adults. The measure assesses older adults' comprehension and processing of everyday printed materials related to IADL performance. The EPT was developed according to psychometric principles of test design, and was validated against older adults' observed performance on selected instrumental tasks of daily living (Diehl, 1991). The measure differs from traditional measures of perceived functional competence by assessing how well subjects actually perform (rather than how well they think they perform) on selected everyday tasks. At the same time, the EPT also emphasizes the cognitive components of everyday task performance (as opposed to physical, social, financial or other components).

Method

Sample

Characteristics

An initial sample of 421 older adults was recruited. Four subjects with incomplete data were dropped from further analysis. Mean age of the participants was 74.88 years (SD = 6.96 years, range = 54-95 years). The subjects were roughly evenly divided into young-old and old-old groups (204 subjects were aged 54-74 years, and 214 subjects were 75 years of age or older). The mean educational level of the sample was 13.15 years (SD = 2.95 years, range = 4-22 years). Participants rated their health from 1 (very good) to 6 (very poor). The mean health self-rating was 2.08 ("good"; SD = 0.96). Average annual income was between $18,000 and $19,999, ranging from under $4,000 to over $50,000.

Recruitment

Subjects were recruited in a stratified fashion from two different sampling frames in the Spring and Fall of 1990. Demographic data to guide the stratified recruitment were obtained from XXXXX (19XX). About half of the sample (n = 202) came from a Florida retirement community; the remaining half of the sample came from senior citizens' apartment complexes located in central Pennsylvania. All subjects were informed that they would receive an honorarium of $30 for their participation. In all cases, site liaisons (administrators) were contacted, and were asked to introduce the study to their residents. Interested residents completed and returned reply cards, indicating their desire to participate. Subject recruitment continued until a desired proportion of subjects in each age-by-education cell was obtained.

Subjects were randomly broken into two groups. One group (MC group) received a multiple choice version of the Everyday Problems Test (EPT). The second group (OE group) received an open-ended (i.e., no answer choices provided) version of the EPT. The MC group contained 213 subjects; the OE group contained 204 subjects (one subject was dropped because of problems during test administration). T-tests were conducted to compare the age, education and health of the two groups; subject groups differed significantly only in education, with the OE group reporting significantly higher educational levels ($p < .05$) (OE group mean = 13.46 years, SD = 2.98; MC group mean = 12.87 years, SD = 2.89). Chi-square tests were conducted to examine group differences in gender and marital status, but no significant group differences were found.

Retest

To assess the test-retest reliability of the EPT, subjects were re-contacted about one year after their initial assessment. A total of 291 individuals (70 males and 221 females) returned to the study. The mean educational level of this returning sample was 12.91 years (S.D. = 2.88 years, range = 6-22 years), and the mean age of the returning sample sub-sample (one year later) was 73.68 years (S.D. = 6.35 years,. range = 59-93 years). T-test comparisons of returning and dropout subjects revealed that returnees were significantly ($p < .05$) younger and less educated, and they reported having lower incomes than non-returning subjects.

Test Development

Item Selection

An initial version of the EPT contained 212 items. The EPT provided subjects with 106 everyday printed stimuli representing seven IADL (Lawton & Brody, 1969) domains: Food Preparation (FOOD), Medication and health behaviors (HLTH), Telephone Use (PHON), Shopping and Consumer Behavior (CONS), Financial Management (FINA), Housekeeping and Laundry ability (HOUS) and Transportation ability (TRAN). Each printed stimulus was associated with two questions about that stimulus. These question pairs were examined following the initial administration of the EPT, and 64 question pairs (128 items) were eliminated. Criteria for the elimination of item pairs included: 1) Low variability. Items pairs with an average proportion correct of .80 or higher (i.e., 20% or less of the sample incorrectly answered the items) were removed; 2) Weak relationship to IADL scale. Item pairs which were negatively or weakly correlated with their hypothesized IADL scale total were also eliminated; and 3) Low face validity. Experts from three professional areas (Physical and Occupational Therapists, Senior Center Directors, and Senior Housing Managers), as well as a panel of older adults, examined the initial set of 106 stimuli and rated them according to their commonality and importance in the everyday lives of older adults (see Diehl, Willis & Schale, 1990, for a complete description of these ratings). This meant that items which had otherwise sound distributional and correlational characteristics, but were rated as unimportant or uncommon, were eliminated. The resulting measure contained 84 items, with six item pairs (12 items) representing each of the seven IADL domains.

Structural analyses

An analysis was conducted to examine whether the seven IADL domains

which the EPT item pairs were written to represent actually provided a good description of the latent constructs underlying the EPT items. To examine this question, a confirmatory factor analyses was conducted using LISREL VI (Joreskog & Sorbom, 1984).

To ascertain whether the same latent structure characterized both the multiple choice and open-ended versions of the measure, a two-groups analysis was conducted. An initial factor analysis specified a 7-factor solution, in which all IADL-domain loadings were freely estimated according to the hypothesized structure (i.e., FOOD item pairs were specified to load only on the FOOD factor, HLTH item pairs on the HLTH factor, and so on). Cross-domain loadings were fixed to zero (e.g., HLTH item pairs were not allowed to load on any factor but HLTH). The analysis was conducted in covariance metric, to permit examination of group equivalence of structure in the original scale metric (Hayduk, 1987). In the initial factor analysis, the psi (PS) matrix of factor intercorrelations was freely estimated, with diagonals elements fixed to 1, so that the scale of the solution was fixed. Diagonal elements of the theta epsilon (TE) matrix of variable residuals were freely estimated; off-diagonal elements (i.e., residual intercorrelations) were fixed to zero. Both the lambda (LY) matrix of factor loadings and the psi matrix of factor intercorrelations were constrained to have equal values in the two groups (multiple choice [MC] and open ended [OE]). The TE values were allowed to vary freely between groups.

The results indicated that the model fit fairly well, but the PSI matrix of factor intercorrelations contained some off-diagonal values in excess of 1.00. Since the confirmatory factor analytic model was attempting to assess the fit of a theoretically-

defined factor structure (Lawton & Brody, 1969), a decision was made not to pursue a more parsimonious first-order factor structure, even though the estimated factor intercorrelation matrix suggested there were very high positive correlations between the hypothesized IADL factors in the EPT. Instead, the investigators opted to fix the values in the PS matrix (i.e., all values in the matrix were predetermined by the investigators). PS matrix values were estimated from the Pearson Product-Moment correlations of the summed IADL scale scores; these correlations were adjusted upward by premultiplying all elements by a constant scalar value which ensured that the maximum off-diagonal correlation would be 0.90. In this second model, LY elements were still constrained to be equal between groups, and TE diagonal elements were again freely estimated, and were allowed to vary between groups.

This second confirmatory factor analysis of the EPT, differentiating between MC and OE groups, suggested that the model fit acceptably: $X^2 = 3227.81$, df = 1680, $p < .001$; goodness-of-fit indices = .622 (MC group) and .792 (OE group). The original covariance matrix was fairly well reproduced: root mean square residual = .066 (MC group) and .050 (OE group). This model demonstrated that the factor loadings and factor intercorrelations could be constrained to be equal across the two groups, suggesting that the multiple choice and open-ended versions of the measure could be represented as structurally similar. Table 1 presents the rescaled (into pseudo-correlation metric, according to Hertzog & Cannon, 1985) factor loadings, factor intercorrelations and standardized unique variances for the EPT 7-factor solution.

The fit of the 7-factor structure of the EPT was also investigated

longitudinally. Based on the accepted two-group invariance model above, data from multiple-choice and open-ended subjects were collapsed into a single covariance matrix. Time One and Time Two EPT item pairs were used to produce the covariance matrix. Based on the Time One solution, the psi factor intercorrelations were again fixed as described above. The results supported configural invariance of the factor solution over time (slight but significant changes in the factor loading values, although the loading pattern remained the same). The obtained fit of the model was $X^2[3353] = 5542.36$, $p < .001$, GFI = .646.

Given the high factor intercorrelations at both occasions, a second-order factor analysis of the EPT was also conducted. At both occasions, a general second-order factor appeared to describe the data well. A metric invariance model (equivalent LY over time, TE allowed to vary) fit the data well ($X^2[76] = 126.38$, GFI = .943).

The results of these measurement model analyses were used to generate composite(factor) scores for further analysis. In some analyses reported below, factor scores from the EPT were generated, using orthonormalized factor loadings as factor scoring coefficients. Factor scores were standardized into T-score metric ($M = 50$, $SD = 10$), to facilitate comparisons across factor scores.

Other measure

Personal data form

Subjects also completed an extensive personal data sheet. In addition to answering questions about their age, education, and occupational history, subjects provided self-ratings of perceived functional competence on eight ordinal

scales representing each of the Instrumental Activity of Daily Living (IADL, Lawton & Brody, 1969) domains: Food Preparation, Medication Use, Telephone Use, Shopping, Financial Management, Housekeeping, Laundry and Transportation Use. Given high proportions of subjects who reported their perceived competence to be unimpaired, responses to each scale were dichotomized (0 = no impairment, 1 = some impairment) and summed.

Results

Analysis of scale level differences

A 2 (Form: MC, OE) X 7 (Domain: FOOD, HLTH, PHON, CONS, FINA, HOUS, TRAN) analysis of covariance was conducted, with Domain as a within-subjects factor, and Form as a between-subjects factor. A General Linear Model solution was used, to adjust for the unbalanced design. Sequential (Type III) Sums of Squares were used in the estimation of F-ratios (SAS Institute, 1985). Covariates were subject age and education, and were included since prior analyses of differences between MC and OE groups indicated slight (but non-significant) group differences in these variables. The dependent variable in this analysis was proportion of correct responses on each of the seven IADL scales of the EPT. Use of the proportional metric permitted examination of scale differences in mean performance level in a common metric, while standardization of each scale would have eliminated these scale mean differences.

Table 2 displays the covariate-adjusted least-squares mean proportions

correct for each EPT scale, in both Forms. Least-squares means are reported to estimate subclass marginal means that would be expected had the design been balanced; these means also hold all covariates at their mean levels (SAS Institute, 1985).

The analysis of covariance revealed a near significant main effect of Domain: $F(6,2478) = 1.97$, $p < .07$. The main effect of Form was non-significant. There was, however, a significant Domain by Form interaction: $F(6,2478) = 17.97$, $p < .0001$.

Planned, non-orthogonal comparisons were conducted, to examine the locus of the Domain by Form interaction. COMCON (Games, et al., 1987) was used to specify a set of seven contrasts; each contrast examined within-domain differences across test forms (eg., comparison of FOOD scale in MC and OE versions). The risk of Type I error was controlled via the Bonferonni alpha adjustment. The open-ended version of the HOUS and HLTH scales had significantly higher means than the multiple choice version ($p < .01$). The multiple choice version of the PHON, FOOD, and CONS scales had significantly higher means than the open-ended versions ($p < .001$). FINA and TRAN scales did not differ significantly between forms. Table 3 presents the obtained t-statistics for each of the planned contrasts.

Measurement characteristics of the EPT.

Internal consistency

Table 4 presents the internal consistency, assessed via Cronbach alpha coefficients, for the EPT. The obtained coefficients were generally moderate-to-high. Coefficients ranged from 0.60 to 0.72 for the scales of the MC version of the EPT, and they ranged from 0.68 to 0.76 for the scales of the OE version. Total scale alphas were 0.93 for the MC test total, and 0.94 for the OE test total.

Test-retest stabilities

Table 5 presents the one-year test-retest correlation coefficients for the EPT and its subscales. The obtained stability coefficients were generally moderate-to-high. Coefficients ranged from 0.46 to 0.72 for the MC version (with a retest stability of .83 for the total measure) and from 0.45 to 0.78 for the OE version (with the test-retest coefficient for the total measure being 0.91).

As part of another investigation, a number of subjects received half the items at their second time of measurement in the opposite format from what they had received at the initial testing (e.g., subjects who had received MC items now received half the items in OE format). For this reason, the retest coefficients presented in Table 5 for the total (84-item) measure could be computed for only a subset of subjects. To compute the test-retest stability for the total sample, only 42 items could be used. As Table 5 also shows, stabilities for this reduced set of items ranged from 0.25 to 0.51 for the MC version, and from 0.52 to 0.68 for the OE version.

To estimate what the test-retest stability of the 84-item version might have been for the total sample, the values obtained for 42-item version were boosted via the Spearman-Brown formula, and are also presented in Table 5. The results are comparable to those obtained with the reduced sample that actually did receive 84

items: Coefficients ranged from 0.40 to 0.68 for the scales of the MC version, and the ranged from 0.66 to 0.81 for the OE version. After the Spearman-Brown correction, total test-retest correlation for the MC version was 0.83, and for the OE version it was 0.93.

Convergent validity.

Convergent validity was examined by comparing subjects who reported above- or below-median levels of perceived functional competence via their self-ratings on the IADL ordinal scales. The median number of self-reported IADL limitations was 1; thus, subjects reporting one-or-fewer limitations were classified as "minimally impaired" on the IADLs; subjects who reported two-or-more limitations were classified as "moderately impaired".

A 2 (IADL Impairment level) X 7 (EPT Domain) analysis of covariance was conducted, with Domain as a within-subjects factor and IADL Impairment as a between subjects factor. Covariates were again subject age and education, and a General Linear Models solution was again used to adjust for the unbalanced design. The dependent variable in this analysis was the proportion of correct responses on each of the seven scales of the EPT.

Table 6 presents the covariate-adjusted mean proportions correct for each scale, separately for each IADL impairment level. For these analyses, 34 subjects who failed to provide complete data for the IADL self-ratings were dropped; thus, these analyses are based on 384 subjects (212 who reported 0 or 1 IADL limitation, and 172 who reported 2 or more IADL limitations). The analysis of

covariance again revealed a near-significant main effect of EPT Domain: $F(6,2280) = 1.85$, $p < .10$, and a significant main effect of IADL Impairment level: $F(1,380) = 4.12$, $p < .05$. For the IADL main effect, subjects with fewer IADL limitations had significantly higher EPT scores (62.9% correct) than subjects with more IADL limitations (59.6% correct). The Domain X IADL Impairment interaction was not significant ($p > .05$): $F(6,2280) = 1.44$.

Planned, non-orthogonal comparisons were conducted, to examine the magnitude of IADL-impairment differences across all seven EPT domains. Seven contrasts were specified to examine within-domain differences between IADL groups. The risk of Type I error was again controlled via the Bonferonni alpha adjustment. Significant IADL group differences were found for 6 of the 7 EPT scales: FOOD ($p < .001$), HLTH ($p < .05$), PHON ($p < .001$), CONS ($p < .01$), HOUS ($p < .05$) and TRAN ($p < .05$). The measure total was also significantly lower for subjects with more IADL limitations ($p < .01$). In all cases, subjects with fewer self-reported IADL limitations had higher EPT scores than subjects with more IADL limitations.

Discriminant validity: Age differences in EPT performance

The performance of young-old (subjects aged 74 and younger) and old-old (subjects aged 75 and older) participants in the study was compared, to see whether EPT scores discriminated between age groups in systematic ways. In general, the mean EPT factor scores of old-old subjects were lower than the scores of young-old subjects. There were variations by scale, however, in the magnitude of the age differences; only the FOOD, HLTH, PHON, and HOUS factor scores were significantly lower for old-old subjects. The standardized total score of older

subjects was also significantly lower than that of younger subjects. Table 7 presents the mean standardized EPT factor scores for the young-old and old-old subjects on each of the EPT domains. Obtained t-statistics for the age group comparisons are also presented.

Discussion

The development of the EPT builds on several decades of research into older adults' competence and perceived competence with the Instrumental Activities of Daily Living (IADLs). At the same time, the EPT differs from the many other measures of older adults' "adaptive behavior" or "functional abilities" (e.g., Beck, 1988; Fillenbaum & Smyer, 1981; Kuriansky & Gurland, 1976; Mahurin DeBettignies, & Pirozzolo, 1991; Spirrison & Pierce, 1992) in at least two ways. First, it is designed for use with community-dwelling populations of older adults, and is designed to profile older adults' strengths and weaknesses in selected domains of everyday activity. It is not designed, explicitly, as a screen for dementia. Second, the EPT emphasizes the cognitive component of everyday competence, focusing on older adults' abilities to comprehend and use the information contained in everyday printed materials. The EPT also emphasizes subjects' objective performance on a set of tasks, rather than relying on subject, proxy, or clinician judgments of competence.

Although the EPT was designed as a measure of the everyday cognitive competence of older adults, it also differs from other assessments of everyday cognition (e.g., Blanchard-Fields, 1986; Camp et al., 1989; Cornelius & Caspi,

1987; Denney & Pearce, 1989; Denney, Tozier, & Schlotthauer, 1992; North & Ulatowska, 1981). The EPT, unlike many other everyday problem-solving measures for older adults, does not present problems of a social or affective nature, and it does not ask subjects to speculate or hypothesize about possible solutions; rather, the EPT focuses on a limited but ecologically valid (Diehl, Willis & Schaie, 1990) subset of tasks for which objective right-or-wrong scores could be derived, and for which there is only one correct answer.

The initial psychometric data presented here indicate that the EPT provides a reliable assessment of its constructs. Based on a sample of older adults selected to represent a broad population of older adults in terms of age and educational distribution, the measure appeared to contain seven highly related domains. Each domain had moderate-to-high internal consistency and test-retest stability was also fairly high over a one-year period. Given the high intercorrelations among domain scores, it is not surprising that the internal consistency and test-retest reliability of the total measure, with its 84 items, was very also high.

The EPT was designed to assess everyday cognitive functioning in seven IADL-type domains. Although seven factors were identified, and remained factorially stable over one year, the high factor intercorrelations and the presence of a strong second-order general factor suggest that multifactorial structure of the EPT may "overfit" the data. This raises an interesting theoretical question: with regard to the cognitive abilities to perform everyday tasks, is cognitive competence general (i.e., the same sorts of skills and knowledge are required in multiple everyday task domains) rather than domain-specific?

Although the common variance defined by the factors appears to be reliable,

both in terms of internal consistency and temporal stability of the identified constructs, it must also be noted that the 7-factor solution for the EPT left substantial residual variance unaccounted for in the item pairs. Other attempted solutions (not reported here) attempted to extract one general factor and smaller local factors simultaneously, but after extracting the first factor, no further factors could be identified. Thus, the unique variance appears to reflect true variance associated with each item pair.

Examining scale level differences, the results showed no significant differences between multiple choice and open-ended versions of the EPT. This means that both in terms of the covariance structure and of the profile of mean difficulty, there were few differences between the two versions of the test. [SHOULD WE DEVELOP THE IMPLICATIONS OF THIS?]

There were significant Domain x Form effects; that is, some scales were slightly easier in the multiple choice version, and others were easier in open-ended format. Examining the magnitude of these differences, however, the sample's proportion correct on a particular scale never differed by more that 10%.

The Domain main effect was also not significant in this study, although the profile of domain difficulty was quite comparable to that reported in large epidemiological studies of IADL self-ratings (Fillenbaum, 1985). There were some noteworthy exceptions, however. Telephone Use, which is typically among the easiest domains in self-rating scales, was the most difficult domain in the EPT. Transportation Use, which is usually one of the most difficult self-rated scales, emerged as one of the easier scales in the EPT. Since the EPT attempts to measure the cognitive capacity for everyday competence, it is likely that self-ratings

are based on multiple judgments (e.g., physical strength, presence of social support), and there should not be a one-to-one correspondence between general assessments of functional competence and a more objective measure of cognitive competence.

The pattern of cross-sectional age differences in this study was fairly typical, and was somewhat supportive of the discriminant validity of the EPT. Although younger subjects performed significantly better than older subjects on most of the EPT subscales, the magnitude of this difference was generally less that 0.3 standard deviation units. This is not very surprising, since all of the participants in the present study were community-dwelling. If the EPT assesses the cognitive skills required to maintain independence in the community, then the absence of large age differences in a community-dwelling sample simply supports the notion that most participants were still cognitively intact.

The present investigation represents a first look at a new measure of older adults' everyday cognitive competence. The psychometric findings in the present paper have shown that the EPT has generally sound measurement characteristics. Future research must establish that 1) EPT performance is predictive of actual everyday competence with the tasks of daily living (Diehl, Willis & Schaie, 1991, found modest correlations between older subjects' performance on the EPT and their performance in a behavioral observation condition); 2) that the EPT discriminates between subjects who are cognitively intact and those who are not (preliminary investigations with demented elders are underway); and 3) that the EPT has utility as a longitudinal predictor of important outcomes for older adults (e.g., maintained functional competence, institutionalization, mortality).

Table 1

Simultaneous two-group complete metric invariance factor structure for the EPT: Seven-factor model

A) Factor loading matrix (LY)

	Factor							Unique variance	
	FOOD	HLTH	PHON	CONS	FINA	HOUS	TRAN	MC	OE
FOOD-PAIR1	.557							.593	.750
FOOD-PAIR2	.568							.678	.677
FOOD-PAIR3	.571							.642	.701
FOOD-PAIR4	.419							.817	.732
FOOD-PAIR5	.544							.731	.671
FOOD-PAIR6	.402							.878	.756
HLTH-PAIR1		.509						.748	.732
HLTH-PAIR2		.572						.668	.678
HLTH-PAIR3		.577						.671	.663
HLTH-PAIR4		.591						.684	.609
HLTH-PAIR5		.560						.680	.693
HLTH-PAIR6		.505						.763	.724
PHON-PAIR1			.445					.780	.820
PHON-PAIR2			.503					.739	.755
PHON-PAIR3			.397					.880	.769
PHON-PAIR4			.361					.868	.872
PHON-PAIR5			.505					.737	.752
PHON-PAIR6			.638					.602	.584
CONS-PAIR1				.465				.779	.787
CONS-PAIR2				.406				.887	.692
CONS-PAIR3				.521				.738	.719
CONS-PAIR4				.379				.852	.861
CONS-PAIR5				.490				.800	.698
CONS-PAIR6				.614				.579	.658
FINA-PAIR1					.519			.765	.683
FINA-PAIR2					.460			.788	.789
FINA-PAIR3					.562			.645	.715
FINA-PAIR4					.451			.784	.808
FINA-PAIR5					.471			.770	.786
FINA-PAIR6					.599			.747	.636

(cont. on next page)

Table 1 (cont.)

	Factor							Unique Variance	
	FOOD	HLTH	PHON	CONS	FINA	HOUS	TRAN	MC	OE
HOUS-PAIR1						.478		.815	.700
HOUS-PAIR2						.474		.757	.792
HOUS-PAIR3						.539		.712	.707
HOUS-PAIR4						.583		.633	.684
HOUS-PAIR5						.435		.758	.823
HOUS-PAIR6						.554		.698	.686
TRAN-PAIR1							.390	.852	.843
TRAN-PAIR2							.581	.692	.625
TRAN-PAIR3							.504	.765	.724
TRAN-PAIR4							.350	.897	.849
TRAN-PAIR5							.572	.667	.679
TRAN-PAIR6							.597	.513	.721

B) Factor correlations (phi)

	Factor						
	FOOD	HLTH	PHON	CONS	FINA	HOUS	TRAN
FOOD	1						
HLTH	.900	1					
PHON	.795	.771	1				
CONS	.841	.882	.824	1			
FINA	.845	.845	.794	.950	1		
HOUS	.870	.894	.819	.933	.887	1	
TRAN	.801	.791	.802	.828	.894	.904	1

Note:
FOOD = Food Preparation, HLTH = Health and Medication Use
PHON = Telephone Use, CONS = Consumerism and Shopping,
FINA = Financial Manage. HOUS = Housekeeping and Laundry,
TRAN = Transportation

Table 2.

Table of covariate-adjusted least-squares means: Domain and Form values

SCALE	MULTIPLE CHOICE	OPEN-ENDED	TOTAL
FOOD	.639	.572	.608
HLTH	.642	.676	.659
PHON	.594	.554	.547
CONS	.615	.535	.576
FINA	.620	.609	.615
HOUS	.578	.635	.606
TRAN	.623	.638	.630
TOTAL	616	.603	.610

Note:
FOOD = Food Preparation, HLTH = Health and Medication Use
PHON = Telephone Use, CONS = Consumerism and Shopping,
FINA = Financial Manage. HOUS = Housekeeping and Laundry,
TRAN = Transportation

Table 3

Comparisons of mean proportion correct: Planned contrasts between test forms, separately for each Domain.

MC - OE CONTRAST[a] FOR SCALE:	OBTAINED DIFFERENCE	OBTAINED T-STATISTIC[b]
FOOD	0.0670	5.531***
HLTH	-0.0340	-2.807**
PHON	0.0400	3.302***
CONS	0.0800	6.604***
FINA	0.0110	0.908
HOUS	-0.0570	-4.705***
TRAN	-0.0150	-1.238

[a] MC = Multiple choice, OE = open-ended

[b] Critical value of t, based on df = 2478 (from within-subjects error term) is 2.696. The Bonferonni adjustment has been performed to control the Type I family-wise error rate. Negative values mean that the open-ended version had a higher mean proportion correct; positive values mean that the multiple choice mean was higher.

Note : * p < .05

FOOD = Food Preparation,
PHON = Telephone Use,
FINA = Financial Manage.
TRAN = Transportation

HLTH = Health and Medication Use
CONS = Consumerism and Shopping,
HOUS = Housekeeping and Laundry,

Table 4.

Internal consistency of the EPT: Cronbach's Alpha

	MULTIPLE CHOICE	OPEN-ENDED
SCALE		
FOOD	.66	.71
HLTH	.72	.76
PHON	.60	.68
CONS	.59	.72
FINA	.68	.70
HOUS	.70	.70
TRAN	.67	.68
TOTAL	.93	.94

Note :

FOOD = Food Preparation,
PHON = Telephone Use,
FINA = Financial Manage.
TRAN = Transportation

HLTH = Health and Medication Use
CONS = Consumerism and Shopping,
HOUS = Housekeeping and Laundry,

Table 5.

Test-retest stability of the EPT

	MULTIPLE CHOICE			OPEN-ENDED		
	84-ITEM	42-ITEM	84-ITEM CORR[b]	84-ITEM	42-ITEM	84-ITEM CORR
	n=53	n=147[a]	n=147	n=54	n=139	n=139
FOOD	.51	.48	.65	.45	.52	.68
HLTH	.46	.50	.67	.55	.56	.66
PHON	.63	.44	.61	.61	.59	.74
CONS	.72	.25	.40	.70	.54	.70
FINA	.55	.45	.62	.78	.62	.77
HOUS	.58	.51	.68	.66	.55	.71
TRAN	.56	.48	.65	.69	.68	.81
TOTAL	.83	.71	.83	.91	.87	.93

[a] All subjects received 42 items in the same format (MC or OE) at both their first and second testing occasions. Only a subset, however, received all 84 items in the same format at both occasions. For this reason, the "true" 84-item test-retest correlation is based on only a limited subgroup of subjects. The 42-item test-retest correlation was computed for the total returning sample.

[b] Corrected 84-item reliabilities represent boosted 42-item coefficients using the Spearman-Brown formula

Note:
FOOD = Food Preparation,
PHON = Telephone Use,
FINA = Financial Manage.
TRAN = Transportation
HLTH = Health and Medication Use,
CONS = Consumerism and Shopping,
HOUS = Housekeeping and Laundry,
EPT = Everyday Problems Test

Table 6.

Covariate-adjusted least-squares means: Domain and IADL group.

SCALE	0-1 IADL LIMITATIONS	2+ IADL LIMITATIONS	TOTAL
FOOD	.636	.581 ***	.611
HLTH	.678	.647 *	.664
PHON	.601	.552 ***	.579
CONS	.597	.563 **	.582
FINA	.623	.616	.620
HOUS	.623	.596 *	.611
TRAN	.646	.620 *	.634
TOTAL	.629	.596 **	.613

***p<.01
*p < .05
+p < .10

Table 7.

Age group differences in EPT performance: Domain factor scores and total score

SCALE	YOUNG-OLD 60-74 (n=204)	OLD-OLD 75-95 (n=213)	T FOR COMPARISON (df = 415)
FOOD	51.15 (9.80)	48.88 (10.09)	2.35*
HLTH	51.17 (9.76)	48.88 (10.12)	2.35*
PHON	51.89 (9.26)	48.19 (10.36)	3.86***
CONS	50.61 (9.42)	49.42 (10.51)	1.22
FINA	50.83 (9.43)	49.20 (10.48)	1.67
HOUS	51.40 (9.39)	48.66 (10.40)	2.82**
TRAN	50.49 (9.82)	49.53 (10.17)	0.98
TOTAL	51.25 (9.53)	48.81 (10.31)	2.51*

Note: Values in parentheses represent standard deviations.
** $p < .01$ *** $p < .001$

Note:
FOOD = Food Preparation,
PHON = Telephone Use,
FINA = Financial Manage.
TRAN = Transportation
HLTH = Health and Medication Use,
CONS = Consumerism and Shopping,
HOUS = Housekeeping and Laundry,
EPT = Everyday Problems Test