Well- and Ill-Defined Measures of Everyday Cognition: Relationship to Older Adults’ Intellectual Ability and Functional Status

Jason C. Allaire
Wayne State University

Michael Marsiske
University of Florida

The present study examined 2 approaches to the measurement of everyday cognition in older adults. Measures differing in the degree of structure offered for solving problems in the domains of medication use, financial management, and food preparation and nutrition were administered to a sample of 130 community-dwelling older adults ranging in age from 60 to 90 (M = 73 years, SD = 7.02 years). Well-defined and ill-defined everyday problem-solving measures, which varied in the amount of means–end-related information provided to participants, were used. The study found that (a) well- and ill-defined measures were moderately interrelated, (b) the 2 approaches were differentially related to basic cognitive abilities, and (c) together the 2 approaches explained over half of the variance in older adults’ everyday instrumental functioning and were in fact better predictors of everyday functioning than traditional psychometric cognitive measures. Discussion focuses on the differential importance of both methods for assessing older adults’ everyday cognitive functioning.

Over the past 20 years, a growing focus of gerontological and psychological research has been on the examination of older adults’ everyday cognitive functioning. The underlying motivation for this field of inquiry has been the concern that psychometric measures of cognition may not appropriately capture real-world cognitive functioning in older adults (Bronfenbrenner, 1979; Conway, 1991; Demming & Pressey, 1957; Denney, 1989; Schaie, 1978; Sinnott, 1989; Sternberg & Wagner, 1986; Wagner, 1986; Willis & Schaie, 1986). The core argument is that traditional laboratory “context-free” measures of cognition may inadequately capture the benefits of experience and familiarity that adults bring to their regular daily tasks. Thus, it has been argued that older adults’ cognitive functioning might more appropriately be assessed with contextually relevant measures assessing their ability to solve problems taken from naturalistic or everyday environments (Allaire & Marsiske, 1999; Berg & Klaczynski, 1996; Blanchard-Fields & Chen, 1996; Cornelius & Caspi, 1987; Wagner, 1986; Willis, 1996; Willis & Schaie, 1986).

The resulting empirical research on adult everyday cognition has been characterized by two somewhat discrepant theoretical approaches, with resulting differences in measurement strategies. The first approach to assessing everyday cognition, which might be labeled the well-defined approach, is grounded in the psychometric and experimental cognitive traditions and proceeds from the assumption that many everyday problems encountered by older adults are highly structured and may often be efficiently solved with a single solution (Allaire & Marsiske, 1999; Hartley, 1989; Hershey & Farrell, 1999; Kirasic, Allen, Dobson, & Binder, 1996; Morrell, Park, & Poon, 1989; West, Crook, & Barron, 1992; Willis, 1991, 1996; Willis & Marsiske, 1991; Willis & Schaie, 1986, 1993). Such well-defined everyday problems are based on a class of traditional problems referred to as well-structured in problem-space theory (Reitman, 1965; Simon, 1973, 1978). As discussed by Eysenck and Keane (1995) as well as by Wood (1983), well-structured problems are those that contain a clear specification of three elements, including the problem situation (initial state), the means or the solution to solve the problem (solution means or operators), and the desired outcome (end state). Often, well-structured problems may be solved by a single correct answer.

Investigators using a well-defined approach have tended to view everyday cognition as being derived from a set of underlying basic abilities. In this view, an amalgam of abilities may be responsible for cognitive performance within everyday contexts rather than any single basic ability (Allaire & Marsiske, 1999; Berry & Irvine, 1986; Marsiske & Willis, 1998; Willis & Marsiske, 1991; Willis & Schaie, 1986, 1993). Thus, any age-related losses in these underlying basic cognitive abilities would be expected to predict resulting deficiencies in everyday cognitive functioning (e.g., Allaire & Marsiske, 1999; Willis & Marsiske, 1991). As age-related decline becomes increasingly “ability-general” in advanced old age (Lindenberger & Reischies, 1999; Schaie, 1996), normative declines in everyday cognition would also be expected in this view. An important caveat to this notion is that compensation may often occur, so the relationship between basic and everyday cognitive declines would never be expected to be perfect.
In contrast with such well-defined problems, some theorists have asserted that problems frequently encountered in daily life are often ill structured as to their goals and their solution. According to problem-space theory (Reitman, 1965; Simon, 1973, 1978), ill-structured or ill-defined problems typically leave at least one of three elements (initial state, solution means, or end state) not clearly specified. Thus, this theoretical perspective would assert that the problems older adults face in their everyday lives frequently tend to offer incomplete or uncertain means of solution and often may be effectively solved with a multitude of potential solutions (Neisser, 1978; Resnick, 1987; Sternberg & Wagner, 1985; Wagner, 1986).

It is important to note that the well- versus ill-defined distinction probably is better characterized as a set of problems varying on a continuum of definedness. Some problems, which leave only the solution means unspecified (such as ours, in this study), would be less ill defined than problems that leave all three elements open (Eysenck & Keane, 1995; Wood, 1983).

In practice, the extant data on ill-defined problem solving have been largely immersed in social psychological and social–cognitive research traditions (e.g., Cantor & Kihlstrom, 1989) and have often focused on the ambiguous and emotion-rich nature of many everyday problems (e.g., P. B. Baltes & Staudinger, 1993; Berg, Meegan, & Klaczynski, 1999; Berg, Strough, Calderone, Sansone, & Weir, 1998; Blanchard-Fields, Camp, & Jahnke, 1995; Cornelius & Caspi, 1987; Denney & Pearce, 1989; Labouvie-Vief, 1985; Strough, Berg, & Sansone, 1996; Wagner, 1986). Many of these researchers have also tended to stress the sociocultural embeddedness of everyday problems, noting that everyday problem solving is shaped by contextual demands, as well as by the goals and perceptions of the individual (e.g., Berg & Calderone, 1994; Berg & Klaczynski, 1996; Klaczynski, 1994). According to problem-space theory, however, social–emotional content is neither a necessary nor sufficient condition for viewing a problem as ill defined. Rather, it is solely the absence of information regarding initial state, solution means, or end state that determines the definedness of a problem.

A common strategy for manipulating this absence of information is to use relatively vague open-ended hypothetical vignettes. A typical example is, “Suppose that an elderly woman needs to go somewhere at night. She cannot see well enough to drive at night and it’s too far to walk. What should she do?” (Denney & Pearce, 1989, p. 439). Because such everyday problems may not include explicit statements about possible solution means or goals and could be effectively solved by a number of possible solutions, they tend to be relatively less well-defined than some of the problems shown in Appendix A. For example, participants might be asked, without any accompanying material, how to avoid foods low in sodium.

The coding of open-ended hypothetical vignettes has been somewhat variable (e.g., Berg & Klaczynski, 1996). In one approach, emphasis is placed on the number or “fluency” of solutions generated to an open-ended everyday problem (Berg et al., 1999; Denney & Pearce, 1989; Marsiske & Willis, 1995; Poon et al., 1992). Many other researchers have used a variety of approaches to examining the style or strategy used to solve everyday problems. For instance, participants’ freely generated responses have been classified into specific strategy styles (Blanchard-Fields et al., 1995; Berg et al., 1998), or participants have been asked to rate the likelihood they would use a given solution corresponding to specific strategies (Blanchard-Fields, Chen, & Norris, 1997; Cornelius & Caspi, 1987; Watson & Blanchard-Fields, 1999). Such approaches have frequently drawn on the coping literature for a taxonomy of strategies (e.g., problem-focused, passive-dependent; Lazarus & Folkman, 1984). Other approaches have focused on classifying open-ended responses into specific cognitive levels based on theories of postformal reasoning (Labouvie-Vief, 1992) or by using highly skilled raters to rate responses on a multidimensional scale (P. B. Baltes & Staudinger, 1993). Although this heterogeneity across investigators might be taken as an indicator of the potential richness of data emerging from hypothetical open-ended vignettes, it also leads to a challenge in being able to generalize the results across studies using different scoring and coding approaches.

From an applied or practical perspective, one way to understand the differences between various everyday problem-solving approaches is to contrast their use in predicting individual differences in important real-world outcomes. Willis and colleagues (Willis, 1991; Willis & Schaie, 1986) have suggested that one outcome of older adults’ everyday cognitive abilities might be the competence of older adults to perform critical tasks of daily living (Willis & Schaie, 1986). Older adults’ functional independence with daily tasks has been widely assessed through the use of self-reported ability within the eight domains of the Instrumental Activities of Daily Living (IADLs; Lawton & Brody, 1969). As Diehl, Willis, and Schaie (1995) noted, research has found that self-reported performance in these domains has substantial predictive utility in identifying older adults at risk for health care utilization (Wolinsky & Johnson, 1991), loss of independence (Wolinsky, Callahan, Fitzgerald, & Johnson, 1992), institutionalization (Branch & Jette, 1982), and mortality (Fillenbaum, 1985). Although a small number of studies have reported a relationship between everyday cognition and measures of adults’ instrumental daily competence (Diehl et al., 1995; Marsiske, 1992; Poon et al., 1992; Willis & Marsiske, 1991), the distinction between well- versus ill-defined everyday problems has not been extended to the question of whether the two approaches differentially predict real-world outcomes.

Relatively few studies have examined this differential prediction question. However, in a few reported studies both well-defined and ill-defined approaches appear to be predictive of IADL competency. For instance, Poon et al. (1992) reported a significant relationship between an ill-defined assessment of everyday cognition and measures of adults’ instrumental daily competence (Diehl et al., 1995; Marsiske, 1992; Poon et al., 1992; Willis & Marsiske, 1991), the distinction between well- versus ill-defined everyday problems has not been extended to the question of whether the two approaches differentially predict real-world outcomes.

From an applied or practical perspective, one way to understand the differences between various everyday problem-solving approaches is to contrast their use in predicting individual differences in important real-world outcomes. Willis and colleagues (Willis, 1991; Willis & Schaie, 1986) have suggested that one outcome of older adults’ everyday cognitive abilities might be the competence of older adults to perform critical tasks of daily living (Willis & Schaie, 1986). Older adults’ functional independence with daily tasks has been widely assessed through the use of self-reported ability within the eight domains of the Instrumental Activities of Daily Living (IADLs; Lawton & Brody, 1969). As Diehl, Willis, and Schaie (1995) noted, research has found that self-reported performance in these domains has substantial predictive utility in identifying older adults at risk for health care utilization (Wolinsky & Johnson, 1991), loss of independence (Wolinsky, Callahan, Fitzgerald, & Johnson, 1992), institutionalization (Branch & Jette, 1982), and mortality (Fillenbaum, 1985). Although a small number of studies have reported a relationship between everyday cognition and measures of adults’ instrumental daily competence (Diehl et al., 1995; Marsiske, 1992; Poon et al., 1992; Willis & Marsiske, 1991), the distinction between well- versus ill-defined everyday problems has not been extended to the question of whether the two approaches differentially predict real-world outcomes.

Relatively few studies have examined this differential prediction question. However, in a few reported studies both well-defined and ill-defined approaches appear to be predictive of IADL competency. For instance, Poon et al. (1992) reported a significant relationship between an ill-defined assessment of everyday cognition (Denney & Palmer, 1981) and self-reported IADL competency. A well-defined measure of everyday cognition has also been found to predict performance on an in-home observation of IADL performance (r = .67; Diehl et al., 1995). Marsiske (1992) included both ill-defined (Cornelius & Caspi, 1987; Denney & Pearce, 1989) and well-defined (Willis & Marsiske, 1993) everyday cognition measures in a single study. In this study, Marsiske (1992) found that self- and proxy-ratings of IADL competency were significantly related to both well-defined and ill-defined measures. Conclusions regarding the differential predictive utility of the two approaches are made difficult, however, because the two approaches often differed in chosen content domains.

Consequently, the goal of the present study was to simultaneously investigate well- and ill-defined approaches to the measurement of everyday cognition in a single sample and their unique and combined abilities to predict adults’ perceived competence...
with tasks of daily living. To facilitate comparisons between more and less well-defined approaches, we designed well- and ill-defined measures that assessed everyday cognitive functioning in the same instrumental everyday domains.

We note here that we view the linkage of everyday problem solving to real-world outcomes, like functional status, as a critical step in advancing this literature (see also Willis, 1991). New measures of everyday cognition must add value in some way for the effort to design and evaluate them to be justifiable. The critical question is whether everyday cognition predicts functional outcomes above and beyond extant cognitive measures. In this context, we are also interested in whether well- and ill-defined measures differed in their predictive ability.

Thus, we conducted the present study to examine three specific questions. First, to what extent were the two approaches to measuring everyday cognition in older adults (well- versus ill-defined) related? It is expected that because the well- and ill-defined measures were designed to assess everyday cognitive functioning within the same everyday domains, the two approaches should be moderately interrelated. Second, how did the well-defined and ill-defined measures relate to traditional psychometric measures of cognition? On the basis of prior research, we expect that the well-defined measures will be more closely related to basic cognitive abilities than the ill-defined measures (Marsiske & Willis, 1995). Third, how did the two measurement approaches relate to indices of everyday functional competence? We anticipate that both measurement approaches will be strongly related to everyday functioning given that they were designed to assess everyday cognitive functioning within instrumental domains of everyday functioning.

**METHOD**

**Sample**

We conducted the present study on a subsample of 130 community-dwelling older adults from the Detroit metropolitan area. This subsample was drawn from a parent sample of participants (N = 174) and represents only those individuals who completed the Open-Ended Everyday Problems Test (OEEPT; Allaire, 1998). The present sample consisted of 24 men and 106 women1 ranging in age from 60 to 90 (M = 73 years, SD = 7.02 years). In addition, the sample was ethnically heterogeneous, with 32% (n = 41) African American, 66% (n = 86) White, 1% (n = 1) Asian/Pacific Islander, and 1% (n = 2) other. Participants reported a mean of 13 years of education (SD = 3.00 years, range = 6–23 years) and a self-reported mean household income of $19,467 (range = $2,000–$50,000+). On a 6-point Likert-type scale ranging from 1 (very good) to 6 (very poor), mean ratings of health and sensory functioning as compared with other same-aged individuals indicated that participants believed their functioning in these domains to be between good and moderately good. Specifically, the mean general physical health rating was 2.30 (SD = 0.97). The mean vision and hearing ratings were 2.57 (SD = 0.89) and 2.56 (SD = 1.06), respectively.

The parent sample, in addition to the 130 participants in this study, included 44 participants who elected not to complete the OEEPT. Comparisons between the current sample and the subsample of participants who completed the OEEPT (n = 130) and those participants in the larger sample who did not (n = 44) indicated that the two groups did not significantly differ in age, t(172) = −1.10, p > .05, education, t(172) = 1.27, p > .05, physical health, t(172) = −0.16, p > .05, vision, t(172) = −0.53, p > .05, and hearing, t(172) = −0.39, p > .05; however, a significant difference in income did emerge, t(172) = 2.16, p < .05, with participants in the complete data sample reporting a higher mean income ($19,468) than participants with incomplete data ($14,905). A detailed description of the parent sample (N = 174) is provided in Allaire and Marsiske (1999). Participants received a small honorarium ($20) for their participation.

**Measures**

The test battery used in this study included traditional psychometric measures of inductive reasoning, declarative memory, and verbal knowledge. We chose these particular basic cognitive abilities for two specific reasons. First, each of the abilities has been studied in several major investigations of psychometric intellectual aging, including the Seattle Longitudinal Study (e.g., Schaie, 1996), the Adult Development and Enrichment Project (e.g., P. B. Baltes & Willis, 1982), and the Berlin Aging Study (e.g., P. B. Baltes & Mayer, 1999), and their cross-sectional and longitudinal trends are well described. Second, previous investigations of everyday cognition have particularly emphasized measures of inductive reasoning, memory, and knowledge as predictors of everyday cognition (Hartley, 1989; Kirasic et al., 1996; Staudeinger, Lopez, & Baltes, 1997; Willis, Jay, Diehl, & Marsiske, 1992; Willis & Marsiske, 1991; Willis & Schaie, 1986).

**Traditional Psychometric Measures**

Because a complete description of the psychometric measures used to assess each of these abilities is provided in Allaire and Marsiske (1999), only a brief overview of these measures is provided here. Inductive reasoning, which is defined as the ability to deduce novel relationships in overlearned material (i.e., letters and numbers), was assessed with the Letter Sets (Ekstrom, French, Harman, & Derman, 1976) and Number Series Tests (Thurstone, 1962). The Verbal Meaning Test (Thurstone, 1962), a multiple-choice measure of recognition vocabulary skills, was included as the psychometric measure indexing participants’ basic knowledgeability. Declarative memory, which is broadly defined as the episodic memory ability to remember information previously encoded, was assessed using the Hopkins Verbal Learning Test (HVLT; Brandt, 1991), a test of participants’ ability to remember a list of 12 common nouns drawn from three semantic categories over three separate trials.

**Everyday Cognition Measures**

In addition to the tests of basic cognitive ability, we included well-defined and ill-defined assessments of everyday cognition. Following Willis and her colleagues (Marsiske & Willis, 1998; Willis & Schaie, 1993), well-defined assessments were marked by everyday problems that are likely to have only a single correct answer, and the inability to produce one correct answer could have serious maladaptive consequences. Ill-defined assessments in this study consisted of everyday problems without clearly specified means and could therefore be effectively solved with a multitude of potential solutions or operators (e.g., Wagner, 1986).

**Well-Defined Everyday Cognition Measures**

Three of the tests from the Everyday Cognition Battery (ECB; Allaire & Marsiske, 1999) were included as representatives of the well-defined

---

1 Preliminary analyses were conducted to ensure that men and woman were similar with respect to the major study variables. Results indicated that the performance of men and women were statistically equivalent on the measures of basic cognitive functioning, t(128) = 0.52, p > .05, on the well-defined measure of everyday cognition, t(128) = 0.35, p > .05, on the solution fluency score, t(128) = 1.35, p > .05, and on the solution quality score, t(128) = 1.16, p > .05, of the ill-defined everyday cognition measures, as well as on overall everyday instrumental functioning, t(128) = 0.38, p > .05.
approach to the assessment of everyday cognition. We focused on three tests from the ECB (i.e., ECB Inductive Reasoning, ECB Declarative Memory, and the ECB Knowledge tests), each of which assessed the everyday manifestation of a single cognitive ability within three instructional domains (i.e., medication use, financial management, and food preparation and nutrition). Participants were asked to answer questions based on real-world stimuli from these domains, and only one correct answer was possible for each question. For the ECB Inductive Reasoning test, participants had to generate answers, and for the ECB Declarative Memory and Knowledge tests, participants had to recognize answers. Solutions from previous analysis with the full parent sample from which this study is drawn (Allaire & Marsiske, 1999) indicated that the internal-consistency reliability of each ECB test was adequate, with Cronbach alpha estimates ranging from .69 to .88. Additional information regarding the content, structure, and psychometric characteristics of the ECB tests are provided in Allaire and Marsiske (1999). In addition, examples of stimuli and items from each ECB test are provided in Allaire and Marsiske and are reproduced in Appendix A.

Ill-Defined Everyday Cognitive Measures

Our relatively more ill-defined approach to everyday cognition was assessed by the OEEPT. This measure was designed to reflect the everyday problem-solving approach advocated by Denney and her colleagues (Denney, 1989; Denney & Pearce, 1989; Denney, Tozzer, & Schlathauer, 1992), and has been used many times in the everyday cognition (e.g., Berg et al., 1999; Blanchard-Fields et al., 1995) and wisdom literatures (e.g., P. B. Baltes & Staudinger, 1993). In the OEEPT, six questions, two in each of the three functional domains, posed a specific everyday problem (e.g., “Your doctor tells you to only eat foods low in fat. What do you do to determine if a particular food at the grocery store is low in fat?”; see Appendix B for a complete listing of the six questions). Participants were asked to generate as many safe and effective solutions to each problem as possible by writing each solution on a separate line. Consequently, the number of solutions generated varied for each problem, within and across participants.

One difficulty, from an empirical perspective, is how to score ill-defined measures such as the OEEPT (e.g., Berg & Klaczynski, 1996). In the present study, drawing on prior research (Denney, 1989; Staudinger, Smith, & Baltes, 1994), we have included two different scoring methods, solution fluency and solution quality.

Solution fluency score. The first scoring procedure, labeled the solution fluency score, derives from the approach used by Denney and her colleagues (Denney, 1989; Denney & Pearce, 1989; Denney et al., 1992) and has subsequently been used in various forms (e.g., number of strategies) by a number of different researchers (e.g., Berg et al., 1999; Blanchard-Fields et al., 1995). In this method the number of solutions determined by raters to be safe and effective is summed, representing a solution fluency score. One assumption might be that the fluency score reflects an individual’s wealth of ideas, with the assumption that the more solutions an individual generates, the more likely it is that individual will succeed in solving the problem.

We acknowledge, however, that the solution fluency score can be theoretically problematic. For instance, Berg et al. (1999) have argued that by virtue of accrued experience in dealing with everyday problems, older adults actually generate fewer solutions because they rely on their prior experience to discount potential solutions (i.e., become increasingly efficient; see also Tordesillas & Chaiken, 1999). It is important to explicate that we chose to include the fluency approach not because we viewed it as an ideal scoring strategy but because (a) it has been used in previous research and we wanted to be able to connect our findings to earlier work, and (b) fluency has an ease and objectivity to it that makes it pragmatically appealing and easily replicable.

The inclusion of safe and effective as solution criteria derives from Denney and Pearce (1989). Again, using an established scoring approach, we were permitted to place our study in the context of earlier work, even as we acknowledge that this approach does not take into account age and individual differences in task representation (Berg et al., 1999; Berg & Calderone, 1994; Berg & Klaczynski, 1996).

Three independent raters scored the solutions for each question either as a 1 if the solution was judged safe and effective or 0 if it was unsafe or ineffective. Raters were instructed to judge a particular solution as safe and effective if it met the following two criteria: (a) if they believed the solution would solve the problem and (b) if the particular solution would not hurt the participant in some way. As an example, for Problem 1 (see Appendix B), a solution rated as effective by all three raters was “Read the nutrition label,” and a solution rated as ineffective by all three raters was “Take a guess.”

Finally, the number of safe and effective solutions was summed across the six questions for each rater, producing a Rater 1 score, a Rater 2 score, and a Rater 3 score. The mean intra-rater correlations for each question ranged from .77 to .96, and for the measure as a whole the mean intra-rater correlation was .95. In addition, estimates of internal consistency, using Cronbach’s alpha coefficient, were calculated across the six questions separately for each of the three raters and were .75 (Rater 1), .73 (Rater 2), and .73 (Rater 3). Collapsing across raters within each of the three domains, the alpha coefficient was .76.

Solution quality score. A second scoring procedure, labeled the solution quality score, used the judgments of another panel of expert raters on the efficacy of each solution generated. This second method derives from one used by P. B. Baltes and Staudinger (e.g., P. B. Baltes & Staudinger, 1993; Staudinger, 1999; Staudinger et al., 1994) in their work on wisdom, although the current study’s approach used a less a priori structure to the rating dimensions than these investigators. Though not unproblematic (Markides, Lee, Ray, & Black, 1993), expert ratings have been reported as a reliable and valid approach in evaluating functioning (e.g., Cohen, Rothman, Poldre, & Ross, 1991).

First, for each of the six questions in the measure, all of the solutions generated by each participant were collapsed into six lists, and redundant solutions were eliminated. Next, a packet containing the resulting six lists of participant solutions was mailed to a convenience sample of 40 professionals who worked in the surrounding community with older adults. These 40 experts consisted of gerontologists primarily working in the fields of social work, nursing, medicine, and administration and who are responsible for making judgments of older adults’ competency on a daily basis. The experts were instructed to rate each solution for its effectiveness in solving the particular problem on a Likert-type scale ranging from 1 least effective to 7 most effective. In addition, they were instructed that effectiveness might encompass dimensions such as efficiency, effort reduction, or use of resources at hand, but they were primarily instructed to rely on their own professional judgments. We intentionally gave loose criteria to the expert raters so that they could rely on their own expertise and conceptualizations of successful problem solving, hoping to maximize the ecological validity of the ratings.

Of the 40 packets distributed, 6 packets were returned with complete data. The six anonymous experts were all women, reflecting the natural gender segregation of these professions, with mean age of 58 years (SD = 7.97, range = 49–71 years) and a mean of 19 years of education (SD = 1.23 years, range = 17–20 years). Four of the six experts identified their ethnic background as White, one was African American, and one expert classified herself as other. Four of the experts were social workers, one was a director of nursing, and the sixth expert was a health care agency administrator. Because the packets were sent out and returned anonymously, we were not able to determine any differences between those experts who participated in the study from those experts who chose not to participate.

Table 1 provides an illustration of the procedure used to create the solution quality score for a single participant. First, the six expert ratings for each solution were assigned to the corresponding solutions generated
by the participant. Second, the ratings for each solution within each problem were averaged separately for each rater producing a mean rating for each of the six problems, which is represented in the rows labeled Expert mean rating. Finally, these expert mean ratings were averaged across the six problems separately for each expert and is represented at the bottom of the table as Overall solution quality score. This procedure was repeated for all participants. Consequently, this scoring procedure produced six overall solution quality scores, corresponding to each of the six expert raters, for each of the 130 participants. The intraclass correlation (Shrout & Fleiss, 1979) across the six questions and the six raters was .81. Estimates of Cronbach’s alpha, calculated for each question of the OEEPT across the six expert raters, were uniformly high with alphas ranging from .81 to .90, indicating that the expert ratings were relatively homogeneous for each of the six OEEPT problems.

**Self-Ratings of Everyday Functioning**

The Older Adults’ Resources and Services modification (Pfeiffer, 1975) of the Lawton and Brody (1969) IADL Scale is an eight-item measure that assesses the perceived level of independence in performing instrumental tasks of daily living such as finance management, medication use, telephone use, shopping, meal preparation, housekeeping, and transportation. Items were recoded so that a response of complete independence in performing a specific daily task was scored as 0 and responses indicating any degree of dependence were scored as 1. Recoded items were then summed to provide an additional indicator of perceived functional ability, with higher scores indicating greater dependence.

The Pfeiffer Functional Activity Scale (Pfeiffer, Kurosaki, Harrah, Chance, & Filos, 1982) was used to assess perceived ability in performing instrumental everyday domains thought to be especially cognitively complex (e.g., financial management, shopping, playing games, food preparation, traveling, keeping appointments, keeping track of current events, and understanding media). Each of the 10 items offered six response alternatives, varying in level of dependence and familiarity of the task, which were recoded so that each item received a score ranging from 1 (complete independence) to 4 (complete dependence). The 10 items were then summed, providing an overall assessment of perceived functional ability. By including both of these measures, it was our intention to broadly assess everyday functioning with a particular emphasis on complex and cognitively loaded everyday tasks.

---

Table 1
Scoring Procedure Used for a Single Participants’ Solution Quality Score

<table>
<thead>
<tr>
<th>Participant 1</th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
<th>Expert 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check the label</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Ask the grocer</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Guess</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Expert mean rating</td>
<td>4.33</td>
<td>4.33</td>
<td>4</td>
<td>3.67</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Problem 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call pharmacy</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Call 911</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Expert mean rating</td>
<td>6.5</td>
<td>5.5</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5.5</td>
</tr>
<tr>
<td>Problem 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call doctor</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Look for it</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Expert mean rating</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Problem 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salt</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Processed meats</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Expert mean rating</td>
<td>5.5</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
<td>5.5</td>
<td>5</td>
</tr>
<tr>
<td>Problem 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call the bank</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Expert mean rating</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Problem 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write a check</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Use credit card</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Expert mean rating</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
<td>7</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>Overall solution quality score</td>
<td>5.81</td>
<td>5.14</td>
<td>5.5</td>
<td>5.20</td>
<td>5.58</td>
<td>5.25</td>
</tr>
</tbody>
</table>
Demographic Characteristics

Each participant completed a personal data questionnaire. The variables of particular interest included in this study were age, gender, ethnicity, income, educational level, and self-evaluations of health, hearing, and vision.

Procedure

The measures in this study were administered as part of a larger study. Two testing sessions took place in the facility (i.e., senior center, church) from which the participants were recruited. During the first 1-hr session, participants completed a consent form and received the demographic questionnaire. Participants returned within 7 days to complete the second session, which lasted approximately 2 hr. During this second session, testing was done in groups ranging in size from 3 to 12 participants. To control for systematic effects associated with practice or fatigue, approximately half the groups received the basic ability measures first, followed by the ECB measures, and the remaining groups received tests in the opposite order.

The measures of everyday cognition were essentially untimed. Participants were given initial periods of 24 and 10 min to complete the ECB Inductive Reasoning test and ECB Knowledge test, respectively. However, participants who needed more time to complete these measures were given as much as they needed during break times or after the formal testing session was over. At the end of the second testing session, participants were given the option to complete a homework packet containing the ill-defined OEEPT. Those participants agreeing to complete the homework were given the option to complete a homework packet containing the well-defined tests, the solution fluency and the solution quality (i.e., composite of Expert Rater 1 through Expert Rater 6) derived from the ill-defined OEEPT. As can be seen in this table, the three well-defined everyday cognition tests (i.e., ECB tests) were all strongly and positively correlated with one another (r = .58 to .71; see also Allaire & Marsiske, 1999). Unlike the well-defined tests, the solution fluency and the solution quality scores derived from the ill-defined OEEPT were not significantly correlated with each other (r = .17).

Turning to the relationships between the two classes of measures, the three ECB tests were significantly and positively related to the solution fluency score (r = .46 to .58) at magnitudes on average only slightly lower than that of the relationships found among the three ECB tests. Although the solution quality score was significantly related to the ECB Inductive Reasoning (r = .23) and ECB Knowledge Tests (r = .25), its relationship with the ECB Declarative Memory test was of substantially lower magnitude (r = .13).

In addition to these bivariate correlations, the relationships among the two measurement approaches were examined at the latent level by using confirmatory factor analysis in LISREL VIII (Jöreskog & Sörbom, 1993). A hypothesized three-factor model was specified that included a factor representing the well-defined approach (Well-Defined ECB factor) and two factors representing the ill-defined approach (i.e., Ill-Defined Solution Fluency and Ill-Defined Solution Quality factors). The Well-Defined ECB factor was identified by the three ECB tests (i.e., ECB Inductive Reasoning, ECB Knowledge, and ECB Declarative Memory). The Ill-Defined Solution Fluency factor consisted of the OEEPT solution fluency scores for Rater 1, Rater 2, and Rater 3. The Ill-Defined Solution Quality factor consisted of the six expert ratings of solution quality derived from the ill-defined OEEPT. The fit of this three-factor model was adequate: χ²(52, N = 130) = 75.04, p < .05, GFI = .92, RMSEA = .06, RMR = .06, CFI = .98, NFI = .95, NNFI = .98, RFI = .93, and IFI = .98. Like the correlations at the observed level, the Well-Defined ECB factor was strongly and positively correlated with the Ill-Defined Solution Fluency factor (r = .65) and considerably less related to the Ill-Defined Solution Quality factor (r = .23), although at the latent level this relationship did reach significance. As expected from the

Table 2
Correlations Between the Well- and Ill-Defined Everyday Cognition Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ECB Inductive Reasoning</td>
<td>—</td>
<td>.70*</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. ECB Knowledge test</td>
<td>.71*</td>
<td>—</td>
<td>.58*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. ECB Declarative Memory test</td>
<td>.58*</td>
<td>.46*</td>
<td>.51*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4. Solution fluency</td>
<td>.23*</td>
<td>.25*</td>
<td>.13</td>
<td>.17</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. ECB = Everyday Cognition Battery. * p < .05.

2 All structural equation models in this study were estimated using the LISREL VIII program (Jöreskog & Sörbom, 1993) and were evaluated using overall fit indices representing estimated goodness of fit (GFI); the root mean square error of approximation (RMSEA); a fit index indicating the discrepancy between the original and reproduced covariance matrix divided by the degrees of freedom and for which values of .05 or lower are indicative of adequate fit; standardized root mean square residual (RMR) (Jöreskog & Sörbom, 1993); comparative fit index (CFI); normed fit index (NFI); nonnormed fit index (NNFI); relative fit index (RFI); and the incremental fit index (IFI) were also examined (Bentler, 1989; Bentler & Bonett, 1980; Bollen, 1989; Marsh, Balla, & McDonald, 1988). In addition to these fit indices, the chi-square estimate was also examined (Akaike, 1987; Carnines & McVey, 1981). The fit of a particular model with most overall fit indices above .90 and a chi-square estimate less than twice the degrees of freedom was considered suggestive of adequate fit.
bivariate correlations, the relationship between the Solution Fluency and Solution Quality factors was the lowest of the three estimated relationships ($r = .19$), but did reach significance at the latent level. 3

Relationship of the Two Approaches to Basic Abilities

Next, the relationships between three basic cognitive abilities (i.e., inductive reasoning, knowledge, and declarative memory) and the ECB, Solution Fluency, and Solution Quality factors were examined. Analyses began with the specification of a six-factor measurement model in which the three everyday cognition factors from the previous section were again estimated (i.e., Well-Defined ECB, Ill-Defined Solution Fluency, and Ill-Defined Solution Quality) along with three factors representing the basic cognitive abilities (i.e., Basic Inductive Reasoning, Basic Knowledge, and Basic Declarative Memory). Specifically, the Basic Inductive Reasoning factor consisted of the Letter Sets and Number Series Tests. The Basic Knowledge factor consisted of two indicators derived from the Verbal Meaning Test, the first representing the sum of the even items (Verbal Meaning 1) and the second reflecting the sum of the odd items (Verbal Meaning 2). The Declarative Memory factor was identified by the three trial scores of the HVLT (Trial 1, Trial 2, and Trial 3). As reported in Allaire and Marsiske (1999), the three basic cognitive ability factors represent the best fitting, most parsimonious solution for these psychometric data, although, given their strong relationships, a second-order general intelligence factor could clearly have been estimated. The fit of this six-factor model was adequate: $\chi^2(138, N = 130) = 177.23, p < .05$, GFI = .89, RMSEA = .05, RMR = .07, CFI = .98, NFI = .92, NNFI = .98, RFI = .90, and IFI = .90. A summary of the standardized parameter estimates for this six-factor model is provided in Tables 3 and 4.

To determine the relative relationship of each basic cognitive ability factor with the three everyday cognitive factors, we specified a structural model, which included direct regression paths from each basic cognitive ability factor (i.e., Basic Inductive Reasoning, Basic Knowledge, and Basic Declarative Memory) to the three everyday problem-solving factors. In addition, the three basic cognitive ability factors were allowed to intercorrelate, as were the ECB and Solution Fluency factors. The fit of this structural model was adequate: $\chi^2(140, N = 130) = 182.05, p < .05$, GFI = .88, RMSEA = .05, RMR = .08, CFI = .98, NFI = .92, NNFI = .98, RFI = .90, and IFI = .90, and did not significantly differ from the measurement model: $\Delta \chi^2(2, N = 130) = 4.82, p < .05$. As can be seen in Figure 1, the Well-Defined ECB factor was strongly related to all three basic ability factors, whereas the Ill-Defined Solution Fluency factor was significantly related to the Basic Knowledge factor. However, none of the three basic cognitive ability factors were significantly related to the Ill-Defined Solution Quality factor.

Use of the Two Approaches in Predicting Everyday Functioning

In this final set of analyses, we examined the relative predictive use of well- and ill-defined assessments of everyday cognition in accounting for individual differences in self-reported instrumental everyday functioning. A four-factor measurement model was specified, which included the Well-Defined ECB and the Ill-Defined Solution Fluency and Ill-Defined Solution Quality factors, as well as an Everyday Functioning factor, which consisted of two indicators representing the Lawton and Brody (1969) IADL Scale and the Pfeffer Functional Activity Scale. The fit of this measurement model was adequate: $\chi^2(72, N = 130) = 108.60, p < .05$, GFI = .90, RMSEA = .06, RMR = .06, CFI = .97, NFI = .93, NNFI = .97, RFI = .91, and IFI = .97. Everyday Functioning was significantly correlated with the Solution Fluency ($r = -.32$) and Solution Quality ($r = -.40$) factors and to a larger extent with the ECB factor ($r = -.69$).

With the factor structure established, a structural model was estimated that included direct paths from the ECB, Solution Fluency, and Solution Quality factors to the Everyday Functioning factor. As in the previous structural model, the ECB and Solution Fluency factors were allowed to correlate. The fit of this model was adequate: $\chi^2(74, N = 130) = 114.92, p < .05$, GFI = .89, RMSEA = .07, RMR = .12, CFI = .97, NFI = .92, NNFI = .96, RFI = .91, and IFI = .97; however, there was a slight loss in fit as compared with the measurement model: $\Delta \chi^2(2, N = 130) = 6.32, p = .04$. As can be seen in Figure 2, the ECB and Solution Quality factors were significant and negative predictors of Everyday Functioning, indicating that poorer performance on the ECB tests and lower open-ended solution quality ratings were related to higher levels of self-reported dependency in Everyday Functioning. Solution Fluency was negatively, but not significantly, related to Everyday Functioning. 4 Together these three factors explained 55% of the variance in Everyday Functioning.

3 More parsimonious two- and one-factor models were also examined as alternatives to the specified three-factor model. The two-factor model consisted of an ECB factor and a second factor (representing all ill-defined problems) that consisted of the three solution fluency scores and the six solution quality scores. The fit of this alternative two-factor model was poor: $\chi^2(54, N = 130) = 415.19, p < .05$, GFI = .57, RMSEA = .23, RMR = .23, CFI = .73, NFI = .70, NNFI = .67, RFI = .64, and IFI = .73. Another two-factor model was also estimated, which allowed the ECB measures and the three solution fluency scores to load on the first factor (representing a general cognitive efficiency score) and the solution quality scores to load on the second. The fit of this two-factor model was also poor: $\chi^2(54, N = 130) = 190.00, p < .05$, GFI = .80, RMSEA = .14, RMR = .11, CFI = .90, NFI = .86, NNFI = .87, RFI = .83, and IFI = .90. Similarly, the fit of a one-factor model, in which all problem-solving scores loaded on a single general factor, was poor: $\chi^2(55, N = 130) = 528.24, p < .05$, GFI = .52, RMSEA = .26, RMR = .25, CFI = .64, NFI = .62, NNFI = .57, RFI = .55, and IFI = .65.

4 In this model, the Solution Fluency factor was not a significant predictor of individual differences in older adults’ everyday functioning. This lack of significance was likely due to multicollinearity between the ECB and Solution Fluency factor. To explore this potential multicollinearity effect, we eliminated the predictive path from the Well-Defined ECB factor to Everyday Functioning. With the removal of this path, the effect of Solution Fluency on Everyday Functioning did reach significance, but together with Solution Quality it explained only 15% of the reliable variance in Everyday Functioning. This finding suggests that the ECB explains both fluency-related variance and additional unique variance in everyday functioning.
Mediation of Basic Abilities

Given the large amount of explained variance in Everyday Functioning, we were interested to see whether the three everyday cognition factors would mediate the basic cognitive ability-related variance in Everyday Functioning. To do so, we first specified a measurement model that included the three basic cognitive ability factors (e.g., Basic Inductive Reasoning, Basic Knowledge, Basic Declarative Memory), the three everyday cognition factors, and the Everyday Functioning factor. The fit of this measurement model was adequate: χ^2(169, N = 130) = 217.67, p < .05, GFI = .87, RMSEA = .05, RMR = .06, CFI = .98, NFI = .91, NNFI = .97, RFI = .89, and IFI = .98. In addition, the Basic Inductive Reasoning, Basic Knowledge, and Basic Declarative Memory factors were all significantly related to the Everyday Functioning factor (r = −.46, −.54, and −.42, respectively).

Next, a mediation model was specified, which included direct paths from the basic cognitive ability factors to each of the everyday cognition factors and from the everyday cognition factors to the Everyday Functioning factor. The fit of the mediation model was also adequate: χ^2(174, N = 130) = 223.67, p < .05, GFI = .87, RMSEA = .05, RMR = .07, CFI = .98, NFI = .91, NNFI = .97, RFI = .89, and IFI = .98, and did not significantly differ from the measurement model: Δχ^2(5, N = 130) = 5.96, p > .05. Next, direct paths from the three cognitive ability factors to Everyday Functioning were added to the mediation model, and the fit was adequate: χ^2(171, N = 130) = 222.50, p < .05, GFI = .87, RMSEA = .05, RMR = .07, CFI = .98, NFI = .91, NNFI = .97, RFI = .88, and IFI = .98. However, even with the addition of these predictive paths, this model did not significantly differ from the previous mediation model: Δχ^2(3, N = 130) = 1.07, p > .05. Thus, the everyday cognition factors (i.e., ECB, Solution Fluency, and Solution Quality) served to significantly mediate all of the basic ability-related variance in Everyday Functioning.

To determine the predictive variance in the Everyday Functioning factor that was unique to, and shared among, everyday cognition and basic cognitive functioning, a communality analysis was performed in the latent space (Hertzog, 1989; Pedhazur, 1982). In this analysis, the ECB, Solution Fluency, and Solution Quality factors were considered as a single predictive block representing everyday cognition, and the three basic ability factors were also considered as a single predictive block representing basic cognition (i.e., Basic Inductive Reasoning, Basic Declarative Memory, and Basic Verbal Knowledge). Estimates of the unique and shared variance components were obtained by allowing all possible combinations of everyday cognition and basic cognitive ability to predict everyday functioning, while maintaining the predictive paths from the basic cognitive ability factors to the everyday cognition factors.

Table 4
Standardized Factor Solution for the Basic Ability and Everyday Cognition Factors (N = 130): Factor Correlations (Phi)

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Basic Inductive Reasoning</td>
<td></td>
<td>.67*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Basic Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Basic Declarative Memory</td>
<td>.52*</td>
<td>.55*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. ECB</td>
<td></td>
<td>.74*</td>
<td>.86*</td>
<td>.66*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Solution Fluency</td>
<td>.48*</td>
<td>.51*</td>
<td>.36*</td>
<td>.65*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Solution Quality</td>
<td>.03</td>
<td>.17</td>
<td>.09</td>
<td>.23*</td>
<td>.19*</td>
<td></td>
</tr>
</tbody>
</table>

Note. ECB = Everyday Cognition Battery.
* p < .05.
As can be seen in Table 5, when the everyday cognition and basic cognitive ability factors were both allowed to predict everyday functioning, they accounted for 59% of the variance. However, the everyday cognition factors uniquely explained 24% of the variance in everyday functioning, whereas basic cognitive ability uniquely explained only 5% of the variance. In addition, over half of the predicted variance in everyday functioning was shared among the everyday cognition and basic cognitive ability factors (30%). Expressed differently, the everyday cognition factors could explain most of the basic ability-related variance in everyday functioning, and substantial unique variance beyond this, but the basic abilities could not explain all of the variance accounted for by everyday cognition.

DISCUSSION

The purpose of the present study was to empirically investigate the relationships between two theoretically derived approaches to the measurement of everyday cognition in older adults. In particular, the study focused on examining three research questions, using more and less well-defined measures of everyday cognition. First, findings from this study suggest that all of our measures of well- and ill-defined everyday cognition are, to some extent, interrelated. A single exception to this general pattern was the relationship between the ill-defined solution quality score and the well-defined ECB Declarative Memory test, which failed to reach significance. Second, the well- and ill-defined measures were differentially related to traditional psychometric cognitive ability tests. Specifically, all of the well-defined ECB factors related strongly and significantly to basic ability factors representing Inductive Reasoning, Verbal Knowledge, and Declarative Memory. The ill-defined factors were not related to the basic cognitive abilities, with the exception of the Solution Fluency factor, which showed a unique relationship with the Basic Knowledge factor. Third, we addressed the question of how well- and ill-defined everyday problems related to everyday functional competence. In addition, we examined the extent to which individual differences in everyday functioning that were related to basic cognitive ability could be mediated by everyday cognition. The results revealed that the Well-Defined ECB factor and the Ill-Defined Solution Quality factor were significant predictors of the everyday functioning composite, explaining more than half of its reliable variance. Follow-up analyses also showed that the everyday cognition factors accounted for most of the basic ability variance in everyday functioning, as well as substantial unique variance beyond that explained by basic abilities. Thus, the key findings from this study were that well- and ill-defined problem-solving factors, as assessed in this study, are not orthogonal and together accounted for substantial and unique variance in a commonly used real-world outcome measure.

In this study, we constrained the social content of the problems included in both the well- and ill-defined measures. In published practice, though well-defined measures have traditionally focused on everyday instrumental domains (e.g., Allaire & Marsiske, 1999; Willis & Marsiske, 1991; Willis & Schaie, 1986, 1993), ill-defined measures typically have dealt with affective or social contextual dilemmas encountered in old age (e.g., P. B. Baltes & Staudinger, 1993; Berg et al., 1998, 1999; Blanchard-Fields et al., 1995; Stough et al., 1996). Even in the domains considered in this study, social context can play an important role. For instance, Margrett (1999) found that older adults considered the activity of preparing meals with others to be a much more social domain than other IADLs (e.g., financial management or medication use). Despite the importance of social-emotional content in everyday problem solving, the problems included in this study were presented without any contextual information regarding the physical location of the problems, the social partners who might be involved, or the interpersonal factors that might guide problem solving. This exclusion of social-emotional problems was a decision in the interest of internal validity, to remove as much of the individual differences in participant preferences and motivations from problem solutions as possible. This tight focus on problems excluding social-emotional content permitted us to more clearly compare the effects.
of the well- and ill-defined problem presentations, without other intervening factors differentiating the two classes of problems.

Future work should seek to determine whether adding back the social–emotional content into everyday problems would aid in accounting for individual differences in everyday functioning. On the one hand, because typical measures of everyday functioning (e.g., Lawton & Brody, 1969) often are themselves fairly well structured and context invariant representing basic competence (M. M. Baltes, Mayr, Borchelt, Maas, & Wilms, 1993; Marsiske, Klumb, & Baltes, 1997), one might argue that measures that capture affective and social context of everyday life would add little to predicting such outcomes. On the other hand, to the extent that ill-defined measures also capture the preferences and motivations that guide real-world decision making, such measures might actually enhance predictions. We take our additional predictive benefit by adding ill-defined solution quality to our models as preliminary support for this notion that including ill-defined measures might enhance predictions. Moreover, if everyday functioning were more broadly defined to include affective and social outcomes, including well-being, it could be that the predictive benefit of ill-defined measures would become even stronger. Indeed, the relative narrowness of our everyday competence outcome is a limitation of the present study. Our self-rating of everyday competence was broader than in many other studies, because it also included items pertaining to cognitively complex activities such as keeping appointments, keeping track of current events, and understanding media (Pfeffer et al., 1982).

In addition to the social–emotional content, we also constrained the everyday domains examined in this study to food preparation, financial management, and medicine use. As explained in Allaire and Marsiske (1999), this decision was governed, in part, by Wolinsky and Johnson’s (1991) finding that such “cognitive activities of daily living” domains might be particularly predictive of functional outcomes like institutionalization and mortality. We did not empirically validate these domains (e.g., factor analyze the scales to confirm cross-measure domain correlations or have judges rate problems to assign them to domains). Rather, in this study, domains were used as a selection criterion to ensure the “everyday-ness” or face validity of the problems used. They were also used to try to ensure that differences between well- and ill-defined problems could not be attributed to different content domains being tested across problem types. It was not our intent to compare functioning across everyday domains. Indeed, several studies have shown (Allaire, 1998; Marsiske & Willis, 1995) that when measures of everyday cognition include multiple domains, a single general latent factor is usually the best representation of between-domain functioning, which suggests that the underlying cognitive processes are fairly similar across domains.

Our chief intent in this study was to contrast well- and ill-defined everyday problems. To build on previous research findings, we therefore used existing measurement approaches (Allaire & Marsiske, 1999; Denney & Pearce, 1989; Marsiske & Willis, 1995). In choosing to draw on these previous approaches, our well- and ill-defined problems differed from each other in two major ways. First, our well-defined problems provided all the information (i.e., initial state, solution means, and end state) needed to solve the problem, whereas our ill-defined problems typically provided or implied an initial state and a goal (e.g., “Your doctor tells you to eat foods low in fat. What do you do?”) but provided no solution means or other information to help solve the problem. Second, our well-defined problems required a single correct answer, whereas our ill-defined problems allowed for multiple correct solutions. From problem-space theory, then, the chief way in which the two classes of problems differed was in terms of the provision of solution means. It must be acknowledged that in using the Denney approach (Denney & Pearce, 1989), our ill-defined problems were not as ill-defined as they might have been if, for example, problem end states had also not been provided. Therefore, it may be theoretically and pragmatically appropriate to view everyday problems as falling along a continuum of definedness. A problem’s exact location along this continuum (e.g., Berg, 1999) is determined by the extent to which the initial state, solution means, and end state of a particular everyday problem is provided or defined.

From these two approaches to everyday problem solving, both used in the research literature but seldom together, we found that both approaches uniquely and significantly predicted real-world outcomes above and beyond psychometric measures of cognition. Establishing the everyday cognition measures as strong predictors of older adults’ everyday competence is an important step in assessing the ecological validity of everyday cognition measures. If everyday problem-solving measures only accounted for the same variance in older adults’ everyday functioning as extant cognitive measures, there would be little justification for adding such measures to batteries containing traditional, well-understood psychometric measures of cognitive functioning. Indeed, our findings clearly suggested that our well- and ill-defined measures, separately and together, explained substantial variance in self-rated everyday functioning beyond traditional cognitive measures. The caveat to this interpretation of our findings is that a broader ability battery might have accounted for more variance in everyday functioning; our current battery included only three cognitive abilities (i.e., inductive reasoning, declarative memory, and knowledge). However, our findings lend preliminary support to the notion that there is “value added” in using everyday cognition measures to predict real-world outcomes. The predictive ability of the well-defined ECB tests and the solution quality score highlights our belief that effective everyday problem solving might be defined in terms of its ability to predict important real-world outcomes. That is, we define effective in terms of positive prediction of a desirable criterion or outcome, and in the context of the present study, higher performance on the ECB tests and higher ratings of solution quality were related to better self-rated everyday functioning.

Of particular interest was our finding that the conclusions drawn from ill-defined problems are critically dependent on scoring ap-
proach. The two scoring strategies used with the ill-defined measure yielded different and independent results. In particular, the solution fluency score was significantly related to basic knowledge, whereas solution quality remained unrelated to basic abilities. Moreover, unlike fluency, solution quality served as a unique predictor of older adults’ self-reported everyday functioning in predictive models that also included well-defined models. In some ways, these findings call into question the use of fluency-based scoring approaches. Although fluency has the advantage of ease and replicability, it offers little unique variance beyond more conventional well-defined measures. Moreover, fluency is conceptually problematic as an indicator of good problem solving. Although one might argue that the more ideas one has available, the more likely one is to select good solutions, a counter possibility could be that expert everyday cognition is characterized by efficiency and the tendency to filter out less-optimal verbosity and inefficiency (e.g., Berg et al., 1999) so that more solutions are not necessarily better.

We acknowledge that our study still leaves important unanswered questions about what strategies or solution types are particularly useful in contributing to positive real-world functioning. Indeed, by using unstructured expert judgments of solution quality, which have the benefit of capturing some of the real-world processes by which everyday competency judgments are made, we gave up the ability to precisely identify the dimensions along which quality judgments were constructed. Speculatively, given that the solution quality score emerged from expert judgments from participants’ written solutions, it is likely that elements of communicative efficiency and person perception (by the raters) may have played a role in those judgments, as might creativity in the generation of unusual and original solutions.

This interpretation is certainly consistent with recent work on wisdom. In a review of their theoretical and empirical work on the topic of wisdom, P. B. Baltes and Staudinger (2000) argued that older adults’ wisdom-related performance reflects the integration of intellectual and personality characteristics. Evidence of this integration is provided by a recent study (Staudinger et al., 1997) in which traditional measures of intelligence uniquely explained very little of the individual differences in older adults’ wisdom-related performance. Instead, measures capturing the intersection of personality and intelligence (e.g., creativity, cognitive style, and social intelligence) accounted for the largest proportion of the explained variance in performance on their wisdom task. Extending this line of argument to the present work, it may be the open-ended, ill-defined problems are also at the interface of cognition and personality, although in the present study we lacked the noncognitive covariate measures needed to test this assertion. Clearly, a critical next step in this literature is to integrate research on the processes, strategies, or dimensions (e.g., Watson & Blanchard-Fields, 1999) of effective problem solving with reliable protocol-based solution quality ratings (expanding on the initial effort in this study) and linking these underlying aspects directly to meaningful real-world outcomes.

Taken as a whole, the results from this study suggest that the current tendency in the research literature to use single measures of everyday problem solving, selected either to reflect a well-defined or ill-defined approach to assessing everyday cognition, may not be the most fruitful. Our findings suggest that the well- and ill-defined measurement approaches are distinct but related, and both may be important in predicting older adults’ everyday functional competence, above and beyond the more context-free measures of cognition typically included in the adult development literature. Expressed differently, these findings can be used to further argue that everyday cognition should not be considered unitary but instead as a multidimensional construct (e.g., Marsiske & Willis, 1995). Thus, the study of everyday cognition will further benefit if a multiple-measurement framework, including both well- and ill-defined measures, is incorporated into future studies.

References


(Appendix follows)
Appendix A

Problem Examples From the Three Everyday Cognition Battery Tests

Example A

1) Miss Braun needs to avoid foods that are high in fat, which can of chili would she be more likely to eat?
2) If she selects Brand B, which categories will she get more of

Example B

1) An expiration or “use by” date on a product means:
   a. the last date the food should be used
   b. last day the product can be expected to be at its peak quality
   c. the date the food was processed or packaged
   d. none of the above

2) When a drug expires:
   a. you should use it up quickly
   b. you should stop taking it
   c. you should see your doctor
   d. you should cut your pills in half to make the prescription last longer

3) If you do not have a checking account, which is probably not a good way to pay your bills?
   a. in person at a bank
   b. send cash in the mail
   c. in person at utility company or mortgage
   d. get a money order

---

**CHILI BRAND A**

| Nutrition Facts | CHILI BRAND B
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Serving Size</strong></td>
<td><strong>Serving Size</strong></td>
</tr>
<tr>
<td>1 cup (236 g)</td>
<td>1 cup (236 g)</td>
</tr>
<tr>
<td><strong>Servings Per Container</strong></td>
<td><strong>Servings Per Container</strong></td>
</tr>
<tr>
<td>about 2</td>
<td>about 2</td>
</tr>
<tr>
<td><strong>Amount Per Serving</strong></td>
<td><strong>Amount Per Serving</strong></td>
</tr>
<tr>
<td>Calories 410</td>
<td>Calories 190</td>
</tr>
<tr>
<td>Calories from Fat 270</td>
<td>Calories from Fat 25</td>
</tr>
<tr>
<td>% Daily Values*</td>
<td>% Daily Values*</td>
</tr>
<tr>
<td>Total Fat 30g</td>
<td>Total Fat 3g</td>
</tr>
<tr>
<td>46%</td>
<td>6%</td>
</tr>
<tr>
<td>Saturated Fat 13</td>
<td>Saturated Fat 1g</td>
</tr>
<tr>
<td>61%</td>
<td>5%</td>
</tr>
<tr>
<td>Cholest. 75mg</td>
<td>Cholest. 75mg</td>
</tr>
<tr>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>Sodium 950mg</td>
<td>Sodium 1250mg</td>
</tr>
<tr>
<td>39%</td>
<td>52%</td>
</tr>
<tr>
<td>Total Carbohydrate 16g</td>
<td>Total Carbohydrate 17g</td>
</tr>
<tr>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>Dietary Fiber 4g</td>
<td>Dietary Fiber 3g</td>
</tr>
<tr>
<td>14%</td>
<td>12%</td>
</tr>
<tr>
<td>Sugars 4g</td>
<td>Sugars 3g</td>
</tr>
<tr>
<td>Protein 20g</td>
<td>Protein 19g</td>
</tr>
<tr>
<td>Vitamin A 26%</td>
<td>Vitamin A 25%</td>
</tr>
<tr>
<td>Vitamin C 0%</td>
<td>Vitamin C 0%</td>
</tr>
<tr>
<td>Calcium 4%</td>
<td>Calcium 3%</td>
</tr>
<tr>
<td>Iron 18%</td>
<td>Iron 15%</td>
</tr>
</tbody>
</table>

* Percent Daily Values are based on a 2,000 calorie diet

---

1) Miss Braun needs to avoid foods that are high in fat, which can of chili would she be more likely to eat?
2) If she selects **Brand B**, which categories will she get more of

---

114 ALLAIRE AND MARSISKE
Example C

Study for one minute:

**DATE OF PRESCRIPTION:** 05-31-96  
**Rx:** 081224

**HOWARD JONES**  
**REFILLS 1**  
**EXPIRES:** 08-31-97

**TAKE 1 TABLET EVERY 6 HOURS (TAKE AT EVEN INTERVALS AROUND THE CLOCK)**

**CAPTOPRIL – 25 MG**  
**90 TABLETS**

**TAKE MEDICATION ON AN EMPTY STOMACH 1 HOUR BEFORE OR 2 TO 3 HOURS AFTER A MEAL UNLESS OTHERWISE DIRECTED BY YOUR DOCTOR**

**THIS DRUG MAY IMPAIR YOUR ABILITY TO DRIVE OR OPERATE MACHINERY, USE CARE UNTIL YOU BECOME FAMILIAR WITH ITS EFFECTS**

Turn the page and answer the questions:

1) How long should Mr. Jones wait to eat a meal after taking a dosage?
   a. he doesn’t have to wait  
   b. 6 hours  
   c. 1 hour  
   d. 2 hours

2) If Mr. Jones takes this medication, what might he be too impaired to do?
   a. write a letter  
   b. drive a car  
   c. talk on the phone  
   d. walk


Appendix B

The Six Problems Included in the Open-Ended Everyday Problems Test (Allaire, 1998)

**Problem 1.** Your doctor tells you to only eat foods low in fat. What do you do to determine if a particular food at the grocery store is low in fat?

**Problem 2.** You accidentally took the wrong combination of medication. What do you do?

**Problem 3.** You have lost your blood pressure medication. What do you do?

**Problem 4.** What foods would you not eat if you were supposed to avoid sodium?

**Problem 5.** You receive your checking account statement in the mail, and there are several checks listed that you don’t remember writing. What do you do?

**Problem 6.** On a Saturday you realize that you don’t have any money in your wallet, and you need to go grocery shopping. What do you do?