

Mental Capacity and Role Taking: A Structural Equations Approach

Author(s): Daniel K. Lapsley and Stephen M. Quintana

Source: *Merrill-Palmer Quarterly*, April 1989, Vol. 35, No. 2 (April 1989), pp. 143-163

Published by: Wayne State University Press

Stable URL: https://www.jstor.org/stable/23086361

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms



Wayne State University Press is collaborating with JSTOR to digitize, preserve and extend access to *Merrill-Palmer Quarterly*

JSTOR

This content downloaded from 71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

Merrill-Palmer Quarterly, Vol. 35, No. 2

Mental Capacity and Role Taking: A Structural Equations Approach

Daniel K. Lapsley
and
Stephen M. Quintana
University of Notre Dame

The determination of the extent to which a latent mental capacity construct can be reliably indicated by standard M-space measures, and whether mental capacity is a significant predictor of role taking development, were the focus of this study. A structural equations approach was used because it accounts for measurement error, generates stable path estimates of predictive relationships, and provides information on the adequacy of the measurement model. Children (n = 99) from grades 1, 3, and 5 responded to two measures of M-space (Counting Span Test and Mr. Cucui) and two measures of role taking (Bystander Cartoons and the Nickle-Dime game). Two structural models were generated. The results indicated that (a) the measurement model was well defined in the two models, (b) M-power was a robust predictor of role taking (Model 1), and (c) the M-power and role taking relation was attenuated when moderator variables were treated as an independent effect (Model 2). The possibility of providing a functional (vs. structural) explanation of social-cognitive development was discussed.

According to neo-Piagetian models of development, the sequential acquisition of various intellectual abilities is constrained by a quantitative mental capacity construct (Case, 1972, 1985; Pascual-Leone, 1970). Mental capacity (or *M-space*) is defined as the maximum number of independent schemes that can be coordinated at any one time. Because mental capacity grows linearly with age (as a result of maturational factors) at the rate of one unit (scheme) every 2 years from the early preoperational stage to maturity, transition between

The authors express their appreciation to the faculty and staff at St. John the Baptist School and St. Matthew School in South Bend, for their kind and patient cooperation in the conduct of this project, and are also grateful to Frank Lorch, Joyce Metti, Lisa Considine, Maher Mouasher, Mike Spretnjak, and Colleen McGinnis for their help in the data collection and scoring, and to Ron Serlin and Charles Scialfa for their statistical suggestions. This project was supported by a grant to Lapsley from the Jesse Jones Research Fund, University of Notre Dame. Correspondence may be sent to Daniel K. Lapsley, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556.

Merrill-Palmer Quarterly, April 1989, Vol. 35, No. 2, pp. 143–163.
Copyright © 1989 by Wayne State University Press, Detroit, Michigan 48202

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

structural developmental substages can be accounted for by successive increases in the size of the "central computing space" that is required to process relevant intellectual schemes. That is, mental capacity sets quantitative limits on the number of schemes that can be coordinated in working memory. Hence, children advance from stage to stage, or progress from nonsolution of a task to solution, when M-power is sufficient to activate task-relevant schemes (Case, 1985).

Considerable evidence in support of an M-operator model of cognitive development has been reported over the years (but see Pascual-Leone, 1978; Romberg & Collis, 1980; Trabasso, 1978; Trabasso & Foellinger, 1978). The evidence is typically one of three kinds: (a) That observed M-values computed from measures of working memory correspond to developmental values of M as hypothesized by Pascual-Leone and Smith (1969) for children of different age groups (e.g., Case, 1972, 1974a, 1974b, 1978; Case & Serlin, 1979; Pascual-Leone, 1970). (b) That different levels of performance on a wide variety of Piagetian logical reasoning tasks are dependent on the attainment of stipulated levels of M-power (Case, 1974b; Pascual-Leone & Smith, 1969; Scardamalia, 1977). And (c) that variations in strategy utilization can be affected by the manipulation of the memory load demanded by a task (Pascual-Leone & Smith, 1969; Scardamalia, 1977).

It is clear that the literature on the validity of mental capacity models of development has centered on cognitive variables and tasks. However, if M-capacity is indeed "a self-regulating developmental mechanism . . . that is universal in a very fundamental sense not only in the sequence of the unfolding stages but also in the rate and timing of development" (Globerson, 1983, p. 229), then it is desirable to know if acquisition in the social-cognitive developmental domain can be similarly accounted for by the mental capacity construct. As Case (1972) points out, developmental capacity constructs should be able to demonstrate cross-task validity, though this has rarely been attempted with social-cognitive variables and tasks (but see, e.g., Chapman, 1981). Although Case (1985, pp. 191–199) attempts to show, by means of a logical analysis, that judgments of fairness and attributions of intelligence do conform to the general pattern of development predicted by his model in the domain of physical cognition, an empirical test of the putative relationship was not reported.

The purpose of the present study was to examine the mental capacity prerequisites of social-cognitive development, using a data analytic strategy that is novel to the M-space literature. In particular, we wanted to determine whether M-power was a significant predictor of

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

role taking development by means of a structural equations analysis (Joreskog, 1974). The selection of role taking as the variable of interest was guided by two considerations. First, there is reason to believe that role taking development may be particularly amenable to a functional explanation in terms of mental capacity. Higgins (1981), for example, argues that advances in role taking ability involve developmental increases in the number of elements or factors (e.g., "perspectives") considered when making a social judgment. The ability to process these factors may be dependent on developmental increases in working memory (Higgins, 1981). Second, role taking is probably the most representative of social-cognitive developmental abilities, in the sense that role taking is foundational for acquisition in other social-cognitive domains, such as moral reasoning (Kohlberg, 1969) and ego development (Selman, 1980; Snarey, Kohlberg, & Noam, 1983). Hence, the demonstration of a "causal" relation between M-power and role taking increases the likelihood that a similar relation obtains as well between M-power and other social-cognitive developmental domains.

A single empirical study (Chapman, 1981) was conducted to account for success on role taking tasks in terms of the Pascual-Leone 1970 M-operator model. Chapman (1981) presented young children with three role taking tasks from the Bystander Cartoons battery developed by Chandler (1973). Mental capacity was assessed with the Backward Digit Span subtest from the WISC. By means of a task analysis, Chapman (1981) determined that role taking success on the Bystander privileged information tasks would be possible only if the child had the mental capacity to coordinate three schemes. These schemes represent the child's understanding that (a) what another person thinks about is the result of knowledge or imagination; (b) knowledge can either be the result of directly perceived knowledge, or indirectly, communicated knowledge; and (c) all relevant information has been presented by the experimenter. On the basis of an "empty cells" binomial test, it was concluded that M-power ($e + 3$ schemes, where e represents a constant executive scheme) is a necessary but not sufficient prerequisite for successful performance on the Bystander Cartoons.

We attempted to provide converging evidence for this conclusion by using multiple measures of both M-power and role taking, and by using a more powerful data analytic strategy. Role taking was assessed by the identical three Bystander Cartoons that were employed by Chapman (1981), and also by the Nickel-Dime game developed by Flavell, Botkin, Fry, Wright, and Jarvis (1968). (See Enright & Lapsley, 1980.)

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

As noted earlier, the Bystander stories are privileged information tasks that assess how well a child can prevent his or her own perspective from intruding on the inference regarding the perspective of a naive bystander. The Nickel-Dime game is a measure of the extent to which a subject can coordinate self-other perspectives in a recursive line of reasoning. The assessment of M-power included the Counting Span Test and Mr. Cucui (Case, 1985; Romberg & Collis, 1980). The former test requires subjects to recall counting operations, and the latter requires memory for spatial location. Hence, in the present study two assessments of role taking (privileged information and recursive reasoning) and two of M-power (counting and spatial location) were included.

The use of multiple assessments allowed the removal of the measurement bias from our constructs. Structural equations is a multivariate technique that identifies spurious relationships and permits the test of models which separate direct from indirect effects and unidirectional from multidirectional effects. Further, this technique allows the segregation of the effects of measurement error though the use of parallel operationalization of latent, unobserved theoretical constructs. Hence, structural equations modeling conveys obvious statistical advantages (a) when it is assumed that constructs are measured with error, (b) when multiple operations are employed, and (c) when the hypothesis under test involves a putatively "causal" relationship, all of which obtain in the present study.

Measurement error has proved to be particularly troublesome in the assessment of M-space. As both Pascual-Leone (1978) and Trabasso (1978) point out, there are many ways to evaluate the M-operator model. Much of the previous research, however, has relied on a limited array of strategies; for example, the strategy of comparing stipulated levels of M-power with performance in some other (usually cognitive) developmental domain, or the strategy of comparing hypothetical and observed values of mental capacity. The model is then typically evaluated by visual inspection of percentages and functions, by nonparametric goodness-of-fit tests, or by contingency table analysis of proportions. In most cases, only a single measure of mental capacity is employed, though there are indications of individual differences and/or measurement error associated with the assessments (Case & Kurland, 1978; Hiebert, 1979; Trabasso, 1978).

In one study, adult (Level 7) subjects did considerably worse than younger subjects on easier M-power tasks (Scardamalia, 1977). Case (1974a) reported that 17% of a 6-year-old sample had a much larger M-space capacity ($> e + 3$) than theoretically would be expected. In another study (Case, 1985, pp. 323–324), the percentage of subjects

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

not performing as predicted across three age groups increased monotonically from 37% to 59% on the Counting Span test, and from 27% to 59% on the Cucui measure.

Finally, in a comprehensive study of M-space measures, Romberg and Collis (1980) found wide variations of M-space scores across different tests of M-space:

> The results of these studies suggested that it is difficult to construct a single measure of M-space which predicts performance on a wide range of tasks. Specific tasks variables, such as stimulus familiarity, may be more important than previously supposed in determining the M-space demand of a particular task. (Romberg & Collis, 1980, p. 56)

The authors suggested, as an alternative, the use of cluster analysis on sets of M-space measures in order to cluster students into groups.

Similarly, it may not always be easy to generate consensus on the quantitative processing demands of particular tasks (Flavell, 1984, p. 201). It is conceivable that different analyses of the same task could arrive at different computations of the memory load and different estimations of the number of "units" required to solve it.

As a result of these considerations, we thought it best to approach the question of the validity of the M-space model from a different direction. In the present study, we conceived M-space and role taking to be latent, unobserved theoretical constructs that can be assessed more or less efficiently by sets of measurement indicators. Hence, we were less concerned with the demonstration of whether success on a role taking task is related to a stipulated level of mental capacity than we were with demonstrating (a) whether multiple measures of mental capacity could define the latent, theoretical M-power construct, and (b) whether M-power, as indicated by a linear combination of mental capacity measures, has a direct relationship to a linear combination of role taking assessments. Structural equations modeling is particularly well-suited to address these questions. In this way, we attempted to provide converging, complementary evidence regarding the universal validity of the mental capacity construct.

METHOD

Subjects

A total of 99 children participated: 37 first graders (23 boys; 14 girls), 31 third graders (13 boys, 18 girls), and 31 fifth graders (13 boys, 18 girls). The average age (and standard deviation) of these groups was 6 years, 3 months (SD = 6 months); 8 years, 3 months (SD = 5 months); and 10 years, 2 months (SD = 5 months), respectively.

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

These age levels were selected to correspond to the theoretical age range that was associated with increments in short term storage space as hypothesized by Case (1985, see Table 14.3) for middle childhood.

Materials

Role taking measures. Two types of role taking tasks, privileged information and recursive reasoning, were employed (Enright & Lapsley, 1980). For the privileged information task, three stories (Airplane, Baker, Sandcastle) from the Chandler 1973 Bystander Cartoons battery were used.

Each story presents an eight-frame cartoon sequence, with two principal characters in each sequence: the protagonist and a late-arriving "bystander." The early frames of the sequence portray events that have an emotional influence on the protagonist. In the Airplane story, for example, a girl is anguished by the departure of her father on an airplane. In the Sandcastle cartoon, a boy's carefully built sandcastle is demolished by a mean-spirited peer. The latter frames of each sequence depict the arrival of a naive bystander who must fathom the reason for the emotional reaction of the protagonist. After describing the course of events in a cartoon sequence, including the relation between antecedent events and the affective reaction of the protagonist, subjects are instructed to return to the point in the sequence where the bystander enters the story, and then relay the rest of the story from the bystander's perspective. The subject has privileged information, of course, about the cause of the protagonist's distress, information that is not available to a late-arriving bystander. To perform successfully on this task, the subject must be cognizant of the naiveté of the bystander and prevent the intrusion of his or her privileged information (the self's perspective) from the inference of the bystander's perspective.

Responses to the three Bystander Cartoons were coded according to a scheme devised by Chapman (1981) which included five categories of responses: (a) The *nonegocentric* status was assigned when the subject stated that the bystander did not know the cause of the protagonist's affective reaction. (b) The *probabilistic* status was assigned if the subject used probabalistic language in attributing privileged information to the bystander. (c) If the subject clearly attributed to the bystander knowledge of events that occurred before the bystander's entrance into the story, that is, attributed privileged information to the bystander, then the subject was scored as *egocentric.* (d) A *noncomprehending* status was reserved for subjects who insisted that they did not know what the bystander thought, or who seemed to for-

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

get the antecedent causes of the protagonist's distress. Finally, (e) subjects were assigned to an *unscorable* category if they gave conflicting accounts of what the bystander could know, or if they simply restated the story.

To accommodate quantitative data analysis, these categories were assigned scores that reflected the developmental adequacy of the response. In order of descending adequacy, the following scale was constructed: 4 = nonegocentric, 3 = probabilistic, 2 = egocentric, 1 = noncomprehending. Unscorable responses were treated as missing data in the data analyses. Data from this assessment were treated in two different ways in the data analysis. For tests of group differences (ANOVA), an overall perspective taking score was derived in accordance with the 1981 Chapman procedure. This score was either the same score assigned to all three stories or else the modal score. Every subject could be assigned a score by this procedure. For the structural equations, analysis scores for individual stories were averaged to yield a single composite score.

The Nickel-Dime game was designed by Flavell et al. (1968) as a measure of role taking activity as opposed to role taking accuracy. This task required children to think recursively about self-other options in a (guessing) game of strategy. Although the basic format of the game is straightforward, the initial version of its administration was somewhat cumbersome because it required the presence of a confederate to serve as the "other." We adapted the game so that the subject was asked to outwit an absent friend instead. In our version of the task, the child was told that she or he was going to play a "tricking" game. He or she was to try to trick a friend, and the friend was going to try to do likewise. The subject was then shown two cups. On the upturned-side of one cup, one nickel was attached. Two nickels (or a dime) were affixed to the upturned side of the second cup. These nickels told the child how many coins were under either cup. The child was shown, with appropriate verbal description, that under the one-nickel cup there was one nickel, and under the two-nickel cup were two nickels, and then told that she or he must choose a cup and remove the coin(s) from under it. But the object of the game was to outwit an opponent (the child's best friend), who would also be asked to pick a cup. The child won if the opponent picked the cup from which the child had removed the coins, and "lost" if the friend picked the cup that still had money under it. The child was told that the friend was aware of the same contingencies and was going to try to outguess him or her. After determining that the child understood how one won and lost, the child was then encouraged to "outsmart" his or her opponent, to win, by figuring out what cup the opponent

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

would probably pick, then taking money out of that one. After a few moments of thinking "read hard," the child was simply asked what cup the opponent would pick and why.

This assessment was scored in the standard way in terms of four categories (Flavell et al., 1968). *Category A* was assigned when a child attributed to the opponent only cognitions and motives regarding the game materials themselves (e.g., picking a cup that yielded the largest monetary advantage), as opposed to cognitions that reflected the opponent's desire to outguess the subject. *Category B* children went on to attribute additional cognitions and motives to the opponent, such that the opponent might opt to alter his or her original selection to counter the subject's initial prediction. Hence, the opponent might be said to pick the one-nickel cup because the opponent would realize that the subject might guess that the "logical" thing to do would be to pick the cup that was more remunerative (the two-nickel cup). *Category C* was assigned if children carried this line of recursive reasoning one or more steps further. *Category O* was assigned if the protocol could not be assigned to any of the preceding categories, which was usually the case when the child could not or would not impute a choice to the opponent, or, after imputing a choice, was unable to offer a rationale for it. Performance on this task was transformed into the quantitative scale of 4 = Category C, 3 = Category B, 2 = Category A, and 1 = Category 0.

M-space measures. The mental capacity measures included the Counting Span Test and Mr. Cucui. The administration and scoring of these measures followed procedures outlined by Romberg and Collis (1980). In the Counting Span Test, children were asked to count arrays of geometric shapes and then recall the number of objects in each array after counting the whole set of arrays. The number of arrays presented to the child increased from trial to trial. M-space was assumed to be equal to the maximum number of arrays that the child could count while maintaining perfect recall (Romberg & Collis, 1980).

The administration of the Counting Span Test included procedures for increasing the familiarity of task instructions, for preventing higher-order "chunking," and for precluding the use of space-saving strategies. To familiarize the child with the stimuli and procedures, several training trials were given before the actual testing. To prevent chunking sequences that employed consecutive numbers, numbers that were either all odd or even were not included. Further, no number appeared twice in a given trial nor occupied the same position on two successive trials. To insure that space-saving strategies could not be used, gray distracter items were included in the array of colored

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

shapes to be counted. Children were also required to count aloud and to point to the shape being counted. Finally, the stimulus cards were presented in succession to prevent rehearsal.

Mr. Cucui requires subjects to store spatial locations. Mr. Cucui is a clown figure that has one or more of its body parts colored (various colors). After viewing the figure for 5 s a blank outline drawing of Mr. Cucui is presented and children are required to point to the parts that are colored. The number of colored parts increases systematically until children reach a point where they can no longer perfectly recall the colored body parts. M-space level is defined as the maximum number of parts that can be perfectly recalled. Case (1985, p. 322) has shown that this measure satisfies the requirements of an M-space test.

To generate mental capacity estimates for the two M-space measures, a scoring procedure devised by Romberg and Collis (1980) was employed. This scoring rule generates a conservative absolute level score that is based on the sum of the total number of correct responses at each level divided by five (the number of items per level).

Procedure

The mental capacity and role taking assessments were administered to each child (individually) on separate days to prevent fatigue effects and to maintain motivation. The choice of the assessment battery to be presented first (M-space or role taking) was determined for each child by a coin flip. The order of administration of the measures within each testing session was also randomly determined. The role taking interviews were audio-recorded for later scoring. Most assessments were completed within 15 to 20 min.

Reliabilities

Both role taking interviews were scored by a trained rater who was not involved in the data collection and who was uninformed about the age of the subject. To establish reliability, a second rater scored 15 randomly selected protocols for either role taking task. The percentage of exact interrater agreement was 87% and 86% for the Bystander Cartoon and the Nickel-Dime game, respectively.

RESULTS

Group Differences

Means and standard deviations for performance on the M-space and role taking measures are reported in Table 1. The actual mean M-

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

Table 1. Means and Standard Deviations for M-space and Role Taking Tasks by Grade

	Grade		
	1	*3*	*5*
	$n = 37$	$n = 31$	$n = 31$
M-Power			
Counting Span			
M	2.45	3.15	3.05
SD	.69	.96	.67
Mr. Cucui			
M	3.05	3.99	4.35
SD	.94	.66	.55
Role Taking			
Bystander Cartoons			
M	2.35	3.40	3.87
SD	1.21	.87	.42
Nickel-Dime			
M	1.59	1.94	2.23
SD	.55	.63	.61

power scores, particularly for the Counting Span Test, correspond rather closely to mean scores on these tests obtained by Case (1985, Table 14.3). As anticipated, scores on the Mr. Cucui measure were generally larger than Counting Span scores at each grade because the former is a relatively easier task (Romberg & Collis, 1980, p. 57). To determine group differences on the M-space and role taking measures, a series of 3(Grade) × 2(Sex) ANOVAs were calculated. The results indicated significant grade main effects for Counting Span, $F(2, 93) = 7.53$, $p < .01$; Mr. Cucui, $F(2, 93) = 26.09$, $p < .01$; the Bystander task, $F(2, 93) = 26.64$, $p < .01$; and the Nickel-Dime task, $F(2, 93) = 9.23$, $p < .05$. Post hoc analysis with the Scheffé procedure showed that the location of the significant differences was between both first and third grade and first and fifth grade for Counting Span, Mr. Cucui, and the Bystander task, and between first and fifth grade for the Nickle-Dime game. No other significant effects were evident.

Structural Equations Analysis

The structural equations technique allows the translation of theoretical models into predicted patterns of relationships and permits the

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

statistical evaluation of how well theoretical models account for observed patterns of relationship among constructs. Hence:

> The structural equation model is used to specify the phenomenon under study in terms of putative cause and effect variables and their indicators. Because each equation in the model represents a causal link rather than a mere empirical association, the structural parameters do not coincide with coefficients of regression among observed variables. (Joreskog & Sorbom, 1984, p. 1)

The fit between a theoretical pattern of prediction among constructs and an observed relationship can be evaluated by a number of goodness-of-fit indices. The chi-square statistic tests the null hypothesis of equivalence between the observed covariances (or correlations) among the measured variables and the covariances reproduced by the structural model. Large (statistically significant) values denote poor fits, small values suggest good fits. The chi-square statistic in structural equations modeling is not robust to sample size and to departures of normality. Hence, additional indices that are not biased by these factors are of interest. The chi-square test is best used when comparing two models, rather than in the assessment of the fit of a single model. The goodness-of-fit index estimates the degree to which the model accounts for the variances and covariances of the data. The adjusted goodness-of-fit index adjusts for degrees of freedom (eliminating bias for models with a small number of degrees of freedom). The root mean square residual is the average of the square of the residuals. The statistical distributions for the latter three indices are unknown, and hence probability values cannot be assigned to the test statistics. However, in a Monte Carlo study by Anderson and Gerbing (1984) it was found that values above .90 for the goodness-of-fit-index, above .80 for the adjusted goodness-of-fit index, and below .05 for the root mean square residual, are indicative of acceptable fits.

The intercorrelations among the M-space and role taking tasks are reported in Table 2. This matrix serves as the data for the structural equations iterations in the LISREL analysis (Joreskog & Sorbom, 1984). The observed correlations reported in Table 2 represent the magnitude of relationship among the measurement indices before measurement error has been removed. Structural equations modeling estimates the magnitude of relations among latent factors after error variance is removed from the observed correlations. The amount of error in each measurement indicator is estimated from the proportion of variance within a single measurement indicator that is not associated with any other measurement indicator in the model. For struc-

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

Table 2. Intercorrelations Among M-space and Role Taking Tasks

		Role Taking		*M-Space*	
	Age	*Nickel-Dime*	*Bystander Cartoons*	*Counting Span*	*Mr. Cucui*
Grade	.958*	.406*	.589*	.311*	.588*
Age		.397*	.612*	.363*	.593*
Nickel-Dime			.264*	.083	.316*
Bystander Cartoons				.255*	.504*
Counting Span					.335*

*$p < .01$.

tural equations modeling to confirm a latent factor structure, within-construct correlations should be higher than between-construct correlations after the disturbance due to error has been removed from all of the correlations. The observed correlations (Table 2) do not demonstrate a pattern of relationship that is strongly indicative of convergent and discriminative validity. As shown later, however, the within-construct correlations will increase relative to the between-construct correlations when measurement error is removed from the intercorrelations, thereby confirming the latent factor structure.

Our strategy for testing and comparing structural equations models was as follows. First, we assessed the factor structure of the data with a full model that includes an over-identified measurement model and a just-identified structural model (see Kenny, 1979, on model identifications). A structural model refers to paths among latent constructs. The measurement model refers to the relation between latent constructs and their measurement indicators. The test of these models is simultaneous, with the evaluation of the full model being dependent on the adequacy of the measurement model. We next evaluated the empirical significance of each path in the structural model (i.e., paths among latent constructs). These paths are evaluated by creating reduced models which differ from the full model by the absence of the to-be-evaluated path. A significant reduction in the goodness-of-fit between the reduced and full models suggests that the presence of the path in the full model is empirically significant. The reduction in the quality of the fit is determined by comparing the chi-square values of the full and reduced models. A significant delta chi-square statistic ($\Delta\chi^2$), which is the statistic that results from the difference between the chi-square values associated with the full and reduced models, indicates that the target path is a significant contributor to the fit of the full model.

In Figure 1 the first of three structural models generated by the analysis is reported. An inspection of path estimates for the measure-

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

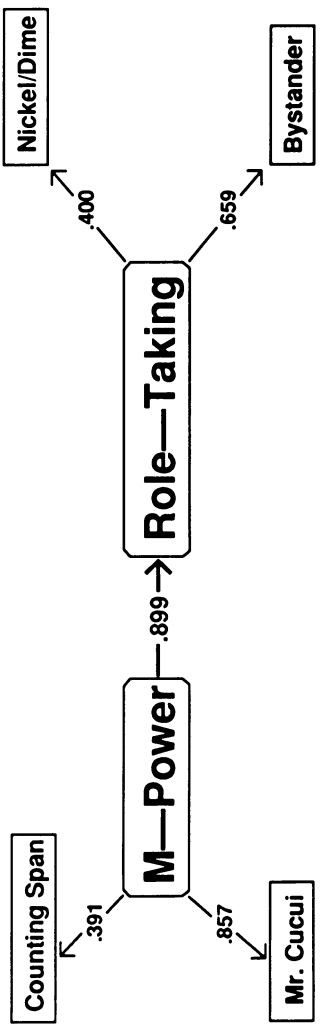


Figure 1. Structural representation of M-power and role taking factors (Model 1). Goodness-of-fit indicators: $\chi^2(1) = 0.61$ ($p = .435$); goodness-of-fit = .997; adjusted goodness-of-fit = .969; root mean square residual = 020.

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

ment model shows that the latent constructs (M-power, role taking) received moderately high factor loadings from their respective indicators. For example , the M-space and role taking measures yielded average path estimates of .582 and .553 for their respective latent constructs (M-power and role taking). Mr. Cucui and the Bystander Cartoons appeared to be more powerful indicators of M-power and role taking, respectively, than were Counting Span and the Nickel-Dime measures, as indicated by path estimate values.

The various goodness-of-fit indices are noted in Figure 1. As can be seen, the model depicted in Figure 1 provided a good fit for the data: $\chi^2(1) = 0.61$ ($p = .435$), goodness-of-fit index = .997, adjusted goodness-of-fit = .969, root mean square residual = .020. The nonsignificant chi-square statistic, and the fact that the values of the latter three indices fall well within ranges that are indicative of acceptable fits, allows the inference that the measurement model is well-defined. That is, the latent constructs (M-space, role taking) are well-defined by their measurement indicators.

Figure 1 also is a report of the path estimate for the structural model. As can be seen, the magnitude of the path estimate between M-power and role taking (.899) suggests that a strong, positive relationship exists between the constructs.

Model 1 was reduced by removing the path between role taking and M-space constructs. The delta chi-square value ($\Delta\chi^2$), comparing the reduced and the full model, was significant, $\Delta\chi^2(1) = 34.22$, $p < .05$. This result indicates that the path is important to the fit of the model and that a strong relationship exists between M-power and role taking.

One threat to the validity of a structural model (or conclusions drawn from the use of any statistical methodology, see Kenny, 1979) is that other factors not included in the model may exert an influence on the factors in the model. Although this threat is never completely eliminated (Serlin & Lapsley, 1985), we attempted to provide stronger support for the pattern of relationships reported in Figure 1 by positing a third summary variable that represents potential, unspecified causal factors. It is conceivable that other maturational and contextual factors may have an influence on the M-power and role taking relationship. Because there are no independent assessments of any particular maturational or contextual factor, we generated a latent moderator variable, indicated by age and grade, to represent a composite or summary variable of unspecified constructs that may be correlated with age and grade.

Figure 2 is the structural representation of this model. As indicated in this figure, this second model provided a viable account of

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

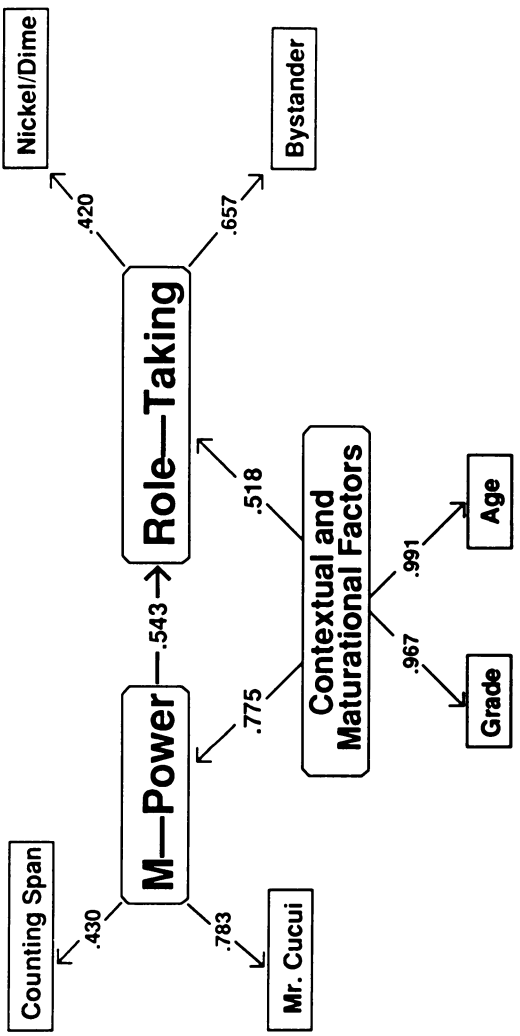


Figure 2. Structural representation of M-power, role taking, and moderator factors (Model 2). Goodness-of-fit indicators: $\chi^2(6) = 5.36$ ($p = .498$); goodness-of-fit = .982; adjusted goodness-of-fit = .936; root mean square residual = .022.

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

the observed data, which is shown by the nonsignificant chi-square statistic, $\chi^2(6) = 5.36$, $p = .498$, and by the fact that the values of the three goodness-of-fit indices are in the appropriate range. It should also be noted that the path estimates between role taking and M-power, and their respective measurement indicators, remained stable across Models 1 and 2, which again suggests that the latent constructs (M-power, role taking) can be reliably identified by the measurement operations used in this study. In Figure 2 is a demonstration also that there was attenuation in the relationship between M-power and role taking in the context of the third latent construct (the moderator variable), which suggests that Model 1 may have overestimated the magnitude of the relationship between M-power and role taking insofar as it did not account for the systematic variance associated with maturational moderator variables.

Having obtained an acceptable fit of the full model for Model 2, three reduced models were generated to assess the empirical significance of causal paths in the full model. The paths of interest included (a) the moderator variable and M-power, (b) the moderator variable and role taking, and (c) M-power and role taking. The results of these tests can be found in Table 3. As can be seen, the link between M-power and the moderator variable was a significant one in Model 2, $\Delta\chi^2 = 46.14$, $p < .05$. The relationship between role taking and the moderator construct may not be an important one, given the nonsignificant chi-square statistic for this comparison. Finally, the M-power and role taking relationship was again empirically significant, $\Delta\chi^2 = 4.40$, $p < .05$.

Two additional analyses were attempted to address possible rival hypotheses. In our first analysis, we paired different indicators with one another to rule out the rival hypothesis that any combination of indicators would form constructs. For example, the Bystander Cartoons measure could be paired with Mr. Cucui, and the Nickel-Dime measure paired with Counting Span, to form a new model. A second model could be constructed by combining Bystander Cartoons with Counting Span, and the Nickel-Dime measure with Mr. Cucui. The tests of these models yielded impermissable path estimates (e.g., correlations substantially above 1.0), which suggested that these two models were not viable representations of the data. Because these models are not subsets of the original three-factor model, delta chi-square statistics could not be computed.

The final test was concerned with whether a single factor "general ability" model, instead of three factors, could adequately represent the observed data. A single factor model can be constructed by positing only one construct and by allowing all of the instruments to

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

Table 3. Model Comparisons

Model	*Path*	$\Delta\chi^2$	*Significance*
1	M-power to role taking	34.22	$p < .05$
2	M-power to role taking	4.40	$p < .05$
	Moderator to role taking	1.32	ns
	Moderator to M-power	46.14	$p < .05$

Note. Each $\Delta\chi^2$ has 1 degree of freedom and a critical value of 3.84.

serve as indicators of this single factor. This model can be statistically compared (via $\Delta\chi^2$) to a full model with three factors because the single factor model is a subset of the three factor model. The resulting comparison was marginally significant, $\Delta\chi^2$ (3) = 7.01, $p < .07$. This finding suggests that the three factor model under consideration cannot be easily reduced to a single general ability model.

DISCUSSION

The purpose of this study was twofold. First, we were interested in determining the extent to which a latent mental capacity construct could be reliably indicated by standard M-space measures. Second, we attempted to determine whether mental capacity was a significant predictor of role taking development. Both aims were addressed by means of a structural equations analysis. This data analytic technique accounted for measurement error, combined observed relationships into latent constructs, generated path estimates of predictive relationships among latent constructs, and provided information on the adequacy of the measurement model. The analysis of the measurement model was a particularly important feature of this study.

The results of this analysis indicate that children's scores on separate measures of mental capacity tend to cluster into a well-defined, latent, M-power construct. This finding is significant, given the problematic nature of M-space assessment that is evident in the literature. The results show that an M-power construct can be reliably indicated by sets of measurement indicators. The Mr. Cucui measure appears to be a more reliable indicator of M-power than Counting Span, although our findings and those of Romberg and Collis (1980) suggest that no one measure is up to the task of assessing mental capacity. In addition, our results show that the Bystander Cartoons outperform the Nickel-Dime game in the assessment of role taking. It appears that role taking is better indicated by the inferential abilities required by privileged information tasks rather than by the recursive reasoning that is required by the Nickel-Dime game.

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

The M-power construct is also a significant predictor of role taking and is seen most clearly in Model 1. The robust prediction of role taking by M-power reported in Figure 1 suggests that any adequate explanation of social-cognitive development must make reference to the mental capacity construct. This finding complements the results obtained by Chapman (1981), who showed that success on the identical Bystander Cartoons used here depended on a stipulated level of M-power ($e + 3$), and gives credence to logical analyses (Case, 1985; Higgins, 1981) that show that various sequences within social-cognitive development are amenable to a mental capacity explanation.

Our results also show that role taking was not directly predicted by the moderator construct, which represented unspecified maturational and contextual factors. It may be the case that, to the extent that maturational and contextual variables do have an influence on role taking development, their influence is mediated by mental capacity. Obviously, our inability to support a direct linkage between moderator variables and role taking may have resulted from insufficient precision of measurement of these variables. That is, age and grade are only proxy indicators of relevant developmental factors. It will be necessary, in future research, to measure moderating developmental variables directly, and to test their effect on role taking independently of the M-space construct.

The rather large goodness-of-fit indices and path estimates may seem surprising, given the moderate size of the correlation coefficients. A number of comments are in order. First, structural equations modeling, as opposed to traditional regression or exploratory factor analyses, depends not on the size of the correlations, per se, but on the consistency of correlations across different constructs. In this respect, the impressive goodness-of-fit indices confirm the predicted latent factor structure. In addition, the path estimates seem large because measurement error has been removed from these estimates. The path coefficients are estimates of the relationship among constructs and indicators when measurement error has been removed. In other words, our results may look "too good" only if the size of observed correlations are inspected and not the factor structure absent measurement error.

A number of caveats should be noted in the interpretation of effects generated by a structural equations methodology. First, the assumption of causality should be granted cautiously, because the correlational feature of the design is not entirely eliminated by this methodology. However, structural modeling does go beyond a correlational approach. Whereas the traditional correlational analysis provides descriptive information on the degree of relationship between

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

two variables, the structural equations technique provides an account of the entire observed pattern of correlations among all of the variables used in the study. That is, all of the observed relationships among indicators and constructs must be well-behaved in accordance with the predicted latent factor structure of the structural model before the measurement model will fit. In addition, the adequacy of this causal explanation must be empirically assessed. That is, different sets of causal predictors can yield different goodness-of-fit indices and parameter estimates. Indeed, we attempted to assess the validity of our models (e.g., Model 2) by comparing them with rival models that represented alternative arrangements of predictors. Furthermore, this data analytic technique also yields information on the convergent and divergent construct validity. Thus, although the design is not entirely devoid of correlational features, neither can the findings that it generates be reduced to mere descriptive correlations.

Another caveat is that the magnitude of the various path estimates could change, perhaps considerably, when other relevant variables are added to the model. The outcome could also change if additional (or different) measurement indicators were used to assess the variables already included in the study. Structural equations modeling is not unlike traditional factor or regression analyses in that statistical outcomes are dependent on the choice of variables and measures that are used in the analysis. The models tested in this study represent a necessary first step in establishing the validity of the M-power construct and in testing causal relations involving the construct. But the models do not exhaust all of the relevant variables, and there are also alternative assessments of both M-space and role taking. Future research will need to consider testing models that include other measures of intellectual ability (e.g., IQ) and social-cognitive development (e.g., socio-moral development).

In sum, the present results show that mental capacity is an important predictor of social role taking development. The significance of this finding is that it supports the extension of neo-Piagetian models to account for acquisition in social-cognitive development. Indeed, because role taking is a foundation or prerequisite for acquisition in other social-cognitive developmental domains, empirical warrant now exists for providing a functional explanation of acquisition in these domains in terms of the mental capacity construct.

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

REFERENCES

ANDERSON, J. C., & GERBING, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49,* 155–173.

CASE, R. (1972.) Validation of a neo-Piagetian mental capacity construct. *Journal of Experimental Child Psychology, 14,* 287–302.

CASE, R. (1974a). Mental strategies, mental capacity, and instruction: A neo-Piagetian investigation. *Journal of Experimental Child Psychology, 18,* 382–397.

CASE, R. (1974b). Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology, 6,* 544–573.

CASE, R. (1978). Intellectual development from birth to adulthood: A neo-Piagetian interpretation. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 33–71). Hillsdale, NJ: Erlbaum.

CASE, R. (1985). *Intellectual development: Birth to adulthood.* Orlando, FL: Academic Press.

CASE, R., & KURLAND, M. (1978). *Construction and validation of a new test of children's M-space.* Unpublished manuscript, University of Toronto (OISE).

CASE, R., & SERLIN, R. (1979). A new model for simulating performance on Pascual-Leone's test of M-space. *Cognitive Psychology, 11,* 308–326.

CHANDLER, M. (1973). Egocentrism and anti-social behavior: The assessment and training of social perspective-taking skills. *Developmental Psychology, 9,* 326–332.

CHAPMAN, M. (1981). Egocentrism and mental capacity: A neo-Piagetian analysis. In J. Meacham & N. Santilli (Eds.), *Social development in youth: Structure and content. Contributions to Human Development* (Vol. 5, pp. 47–63). Basel, Switzerland: S. Karger.

ENRIGHT, R., & LAPSLEY, D. (1980). Social role taking: A review of the constructs, measures, and measurement properties. *Review of Educational Research, 50,* 647-674.

FLAVELL, J. (1984). Discussion. In R. Sternberg (Ed.) *Mechanisms of cognitive development* (pp. 187–210). New York: Freeman.

FLAVELL, J., BOTKIN, P., FRY, C., WRIGHT, S., & JARVIS, P. (1968). *The development of role taking and communication skills in children.* New York: Wiley.

GLOBERSON, T. (1983). Mental capacity and cognitive functioning: Development and social class differences. *Developmental Psychology, 19,* 225–230.

HIEBERT, J. (1979). *The effect of cognitive development on first grade children's ability to learn linear measurement concepts.* (Tech. Rep. No. 506). Madison: University of Madison, Wisconsin Research and Development Center for Individualized Schooling.

HIGGINS, E. T. (1981). Role taking and social judgement: Alternative developmental perspectives and processes. In J. Flavell & C. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 119–153). Cambridge: Cambridge University Press.

JORESKOG, K. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D. Krantz, R. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Measurement, psychophysics, and neural information processing* (pp. 1–56). San Francisco: Freeman.

JORESKOG, K., & SORBOM, D. (1984). *LISREL* VI (3rd ed.). Mooresville, IN: Scientific Software.

KENNY, D. A. (1979). *Correlation and causality,* New York: Wiley.

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms

KOHLBERG, L. (1969). Stage and sequence: The cognitive developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.

PASCUAL-LEONE, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologia, 32,* 301–345.

PASCUAL-LEONE, J. (1978). Compounds, confounds, and models in developmental psychology: A reply to Trabasso and Foellinger. *Journal of Experimental Child Psychology, 26,* 18–40.

PASCUAL-LEONE, J., & SMITH, J. (1969). The encoding and decoding of symbols by children: A new experimental paradigm and a neo-Piagetian model. *Journal of Experimental Child Psychology, 8,* 328–355.

ROMBERG, T., & COLLIS, K. (1980). *The assessment of children's M-space.* (Tech. Rep. No. 540). Madison: University of Wisconsin-Madison, Wisconsin Research and Development Center for Individualized Schooling.

SCARDAMALIA, M. (1977). Information-processing capacity and the problem of horizontal decalage: A demonstration using combinational reasoning tasks. *Child Development, 48,* 28–37.

SELMAN, R. (1980). *The growth of interpersonal understanding.* New York: Academic Press.

SERLIN, R., & LAPSLEY, D. (1985). Rationality in psychological research: The good-enough principle. *American Psychologist, 40,* 73–83.

SNAREY, J., KOHLBERG, L., & NOAM, G. (1983). Ego development in perspective: Structural stage, functional phase, and cultural age-period models. *Developmental Psychology, 3,* 303–338.

TRABASSO, T. (1978). On the estimation of parameters and the evaluation of a mathematical model: A reply to Pascual-Leone. *Journal of Experimental Child Psychology, 26,* 41–45.

TRABASSO, T., & FOELLINGER, D. (1978). Information processing capacity in children: A test of Pascual-Leone's model. *Journal of Experimental Child Psychology, 25,* 1–17.

This content downloaded from
71.239.98.242 on Wed, 14 Oct 2026 01:49:38 UTC
All use subject to https://about.jstor.org/terms