The Construct Validity of Memory Span as a Measure of Intelligence

Bruce L. Bachelder
Independent Practice, Retired
Morganton, NC

A presentation at the
Annual North Carolina Cognition Conference
Chapel Hill, NC
March 2007

A Clash of Theories

- Working Memory Capacity (WMC) theory in the cognitive science tradition
- Span theory in the behavioral tradition

Working Memory Capacity Theory: Measures of working memory are the central measures of developmental and individual differences

Span Theory: “Simple memory span” measures are central

Simple span test: Subjects hear a series of digits or words and try to say them back perfectly. The score is a measure of the largest number of items which can be done perfectly.

Working memory test: Subjects also hear and report series of words or digits, but this is done concurrently with another task, such as reading or mental arithmetic.

WMC Theory vs. Span Theory

- WMC measures “correlate with a wide range of other cognitive measures. . . . This list is particularly impressive given the notable lack of such a relationship with simple span measures of temporary memory” (Engle & Kane, 2004, p. 153)
- This statement is at odds with decades of data

My Main Points Today

- The relationship of memory span to general intelligence is actually greater than is generally believed (Arthur R. Jensen, 1970, pp. 71-74)
- As measures of intelligence, span tests have excellent construct validity

References

- Bachelder & Denny, 1977 a,b (span theory)
- Bachelder, 1977 (span & language)
- Blankenship, 1938 (memory span)
- Jensen, 1964, 1970 (span and intelligence)

Two types of Validity

- *Face validity* is a judgment based on appearance, domain content, and tradition
- *Construct validity* is an empirical matter, data evidence that test scores behave as theory demands

Construct Validity

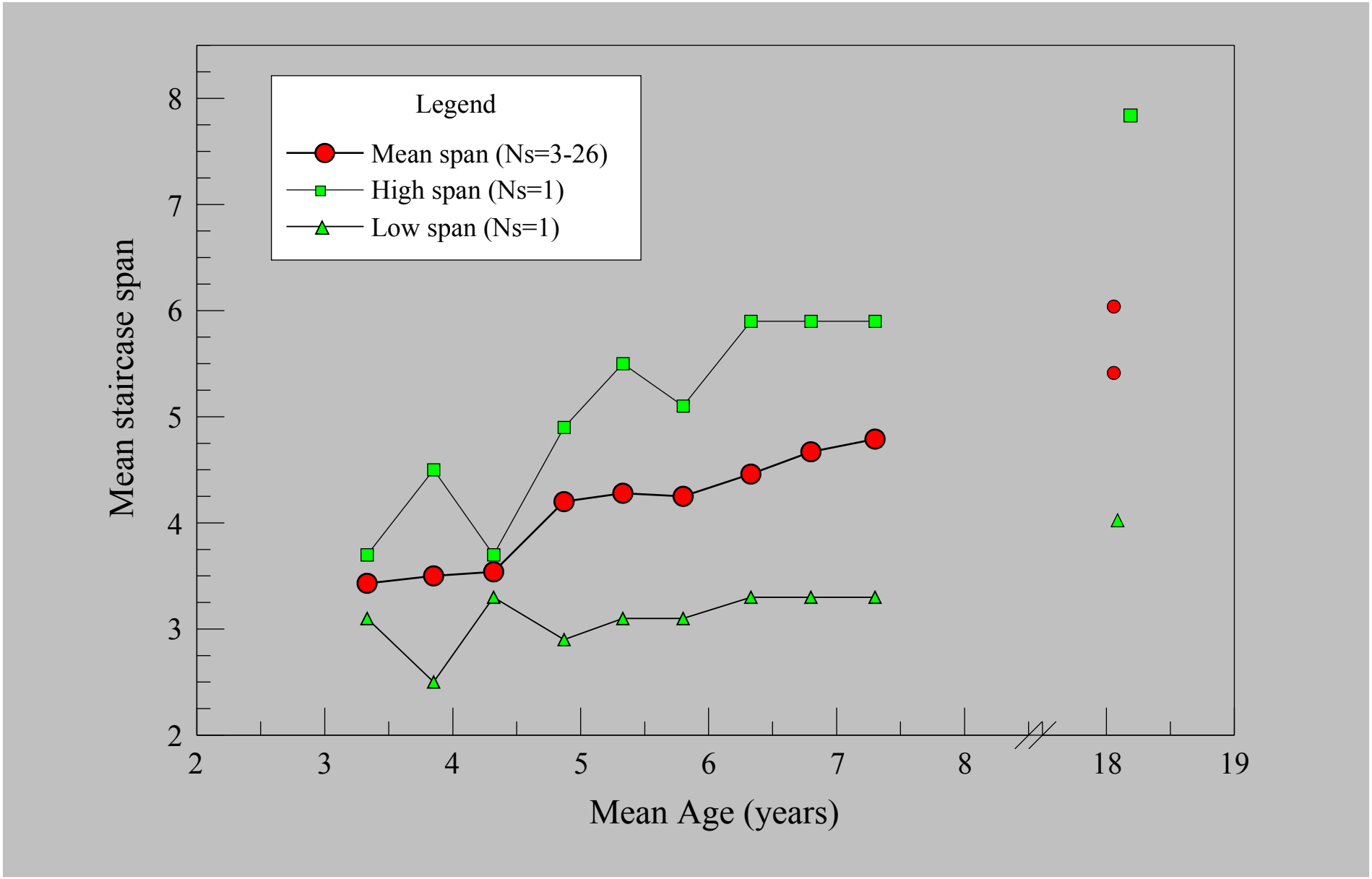
- If it walks like a duck
- If it quacks like a duck
- Maybe it is a duck
- Memory span scores “walk and quack” like intelligence

The stability of span scores

- Test-retest correlations of the standard staircase span test run about .80
- Reliability can be taken about as high as one would want, to the mid .90s, simply by increasing the number of test trials (Arthur R. Jensen, 1970, p. 72)

Age and Individual Differences

- Spans increase during the developmental period and asymptote in early adulthood
- At each age there are stable individual differences
- Variance appears to vary directly with age during the developmental period


Legend
Mean span (Ns=3-26)
High span (Ns=1)
Low span (Ns=1)
Mean staircase span
Mean Age (years)
2
3
4
5
6
7
8
18
19

Robust IDs & Developmental Effects

- No gender effect
- Little or no improvement with practice or training
- Little or no race effect
- Little or no culture effect
- Comparatively minor materials effects

Correlation with Academic Achievement

- Students at the bottom of their class by teacher judgment have smaller spans (Jacobs, 1887)
- DS is the best predictor in a comprehensive battery of tests for predicting 1st grade performance for “high risk” kids (Serwer, Shapiro, & Shapiro, 1972)

Correlation with IQ

- 0.79, DS and IQ, Bachelder (1970/1971)
- 0.66, WS and IQ, Bachelder (1976)
- 0.63 (.80 corrected for attenuation), Wechsler's DS and *g*, (Jensen, 1970)
- .80 (corrected for attenuation), MA and DS (Butterfield, 1968, corrected by Bachelder & Denny, 1977)

More Correlations with IQ

- .60 - .70 (corrected for attenuation), WISC Digit Span and FSIQ (minus DS)
- .75 (corrected for attenuation), WAIS DS with FSIQ (minus DS)
- .62, Digits, Stanford Binet, age 2 1/2
- Span tests load .50 on fluid intelligence, .00 on crystallized intelligence (Horn, 1968)
-

Correlation with Cognitive Ability

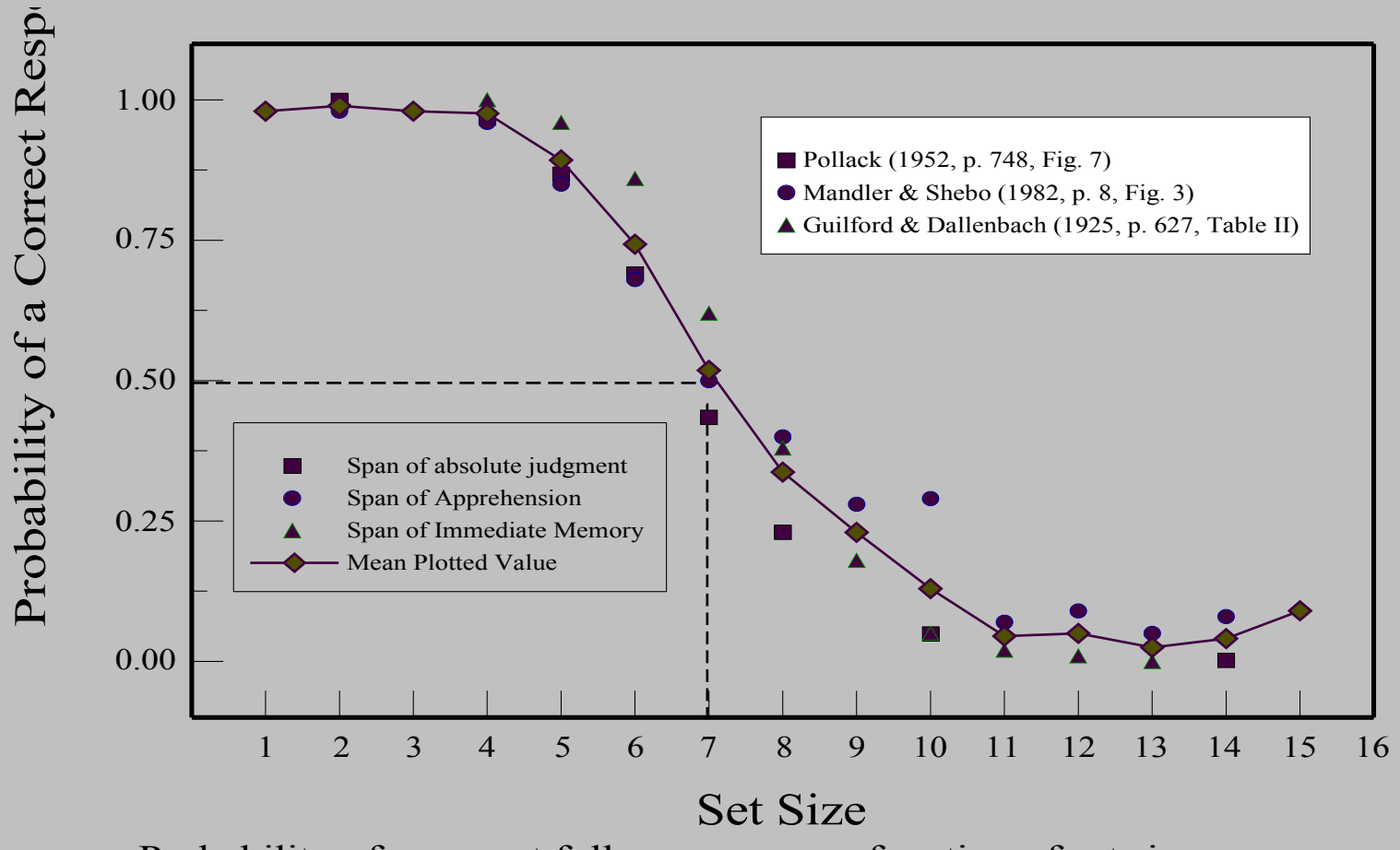
- All mental tests correlate with all other mental tests, the *positive manifold,* Jensen, 1986
- College students have higher span scores than average adults (Bachelder, 1976)
- .87, 6 STM tests with *g* from 4 tests (motor speed, vocabulary, arithmetic, form board) (Garrett, Bryan, & Perl, 1935)

Covariation with Language

- Sentence imitation (Bachelder, 1977)
- Symbol-board span test
- Sentence generation

Span Ability Transcends Specific Test Materials

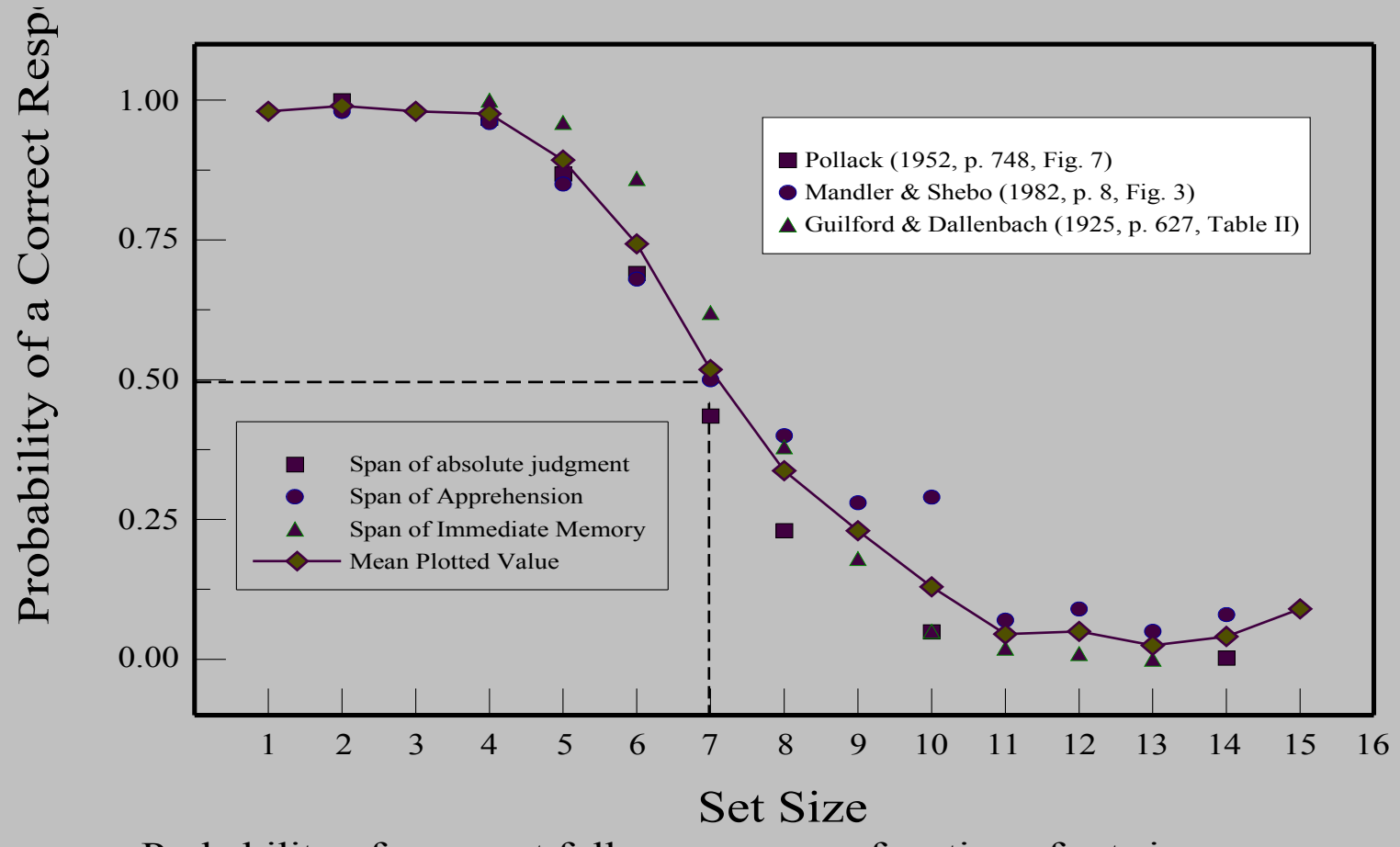
- Digits, words, color samples, all lead to similar span values
- Span tests are highly intercorrelated
- A factor analysis produces a strong span factor (Brener, 1940)
- Symbol board span tests yield valid scores



Probability of a correct full response as a function of set size.

Span Ability Interacts with Task Difficulty

- Brighter individuals handle difficult or complex tasks as well as simple tasks
- Developmentally younger or intellectually slower individuals cope well only with simpler tasks



Probability of a correct full response as a function of set size.

Better Construct Validity than the IQ

- Little or no black-white difference
- Culture-free (Jensen, 1964)
- No Flynn effect
- One simple test is usable with nearly all Ss
- Individualized face-valid tests can be developed readily as needed

Better Construct Validity than the IQ

- Not vulnerable to the criticism that it was constructed to behave like intelligence
- Easily interpretable in basic scientific terms common to cognitivism, behaviorism, and lay theory alike, namely, S, R, and Task

How Could this Happen?

- Some psychometric considerations: constricted ability range, low reliabilities, diverse test methods
- Theory is the tail wagging the dog
- It is the nature of “normal science,” (Kuhn, 1972) to ignore, resist, and explain away data contrary to expectation, and focus on research to support expectations

WMC Theory vs. Span Theory

- Simple span measures "tell us primarily about domain-specific rehearsal processes, such as inner speech, and domain-specific knowledge, for example, pertaining to word meanings or the recognition of salient digit patterns" (Engle & Kane, 2004, p. 150).
- Simple span tests measure *span ability* which transcends content and underlies *intelligence* (Bachelder & Denny, 1977a, b)