Non-Linear Growth Models as Psychometric Models: A Second-Order Growth Curve for Measuring Potential

Dan McNeish
Utrecht University

Denis Dumas
University of Maryland

Developed vs. Developing Constructs

- In the context of psychological measurement, there are two general types of constructs that can be measured:
  - Develop<u>ed</u> constructs are static and represent extant skills at the time of measurement.
  - Develop<u>ing</u> constructs are dynamic and represent learning and change that may occur in the future.
- When measuring psychological constructs, there is often an interest not only in how much of the construct a person has when measured (a develop<u>ed</u> construct) but also how much they could develop in the future (a develop<u>ing</u> construct)

Developed vs. Developing Constructs

- The current state of psychological measurement often does not differentiate between these two types of constructs

- For example, the ACT and the SAT are interested more in future academic performance (develop<u>ing</u>) than current academic performance in high school (develop<u>ed</u>)

- That is, measures for develop<u>ed</u> constructs are utilized in circumstances where the construct of interest is actually develop<u>ing</u>

Trouble with Developing Constructs

- This discrepancy is understandable as develop<u>ing</u> constructs are far more difficult to measure than develop<u>ed</u> constructs

- Traditional CFA or IRT models are not very helpful

- A long line of research dating back to the 1950s and 1960s used multiple testing occasions, intermingled with strategy instruction, to tap developing constructs (e.g., Feuerstein, Rand, & Hoffman, 1979)
  - One major goal of DA is improving validity for test-takers from non-dominant groups (e.g., SES, Race, Gender)

- This technique has been termed: *Dynamic Assessment (DA)*

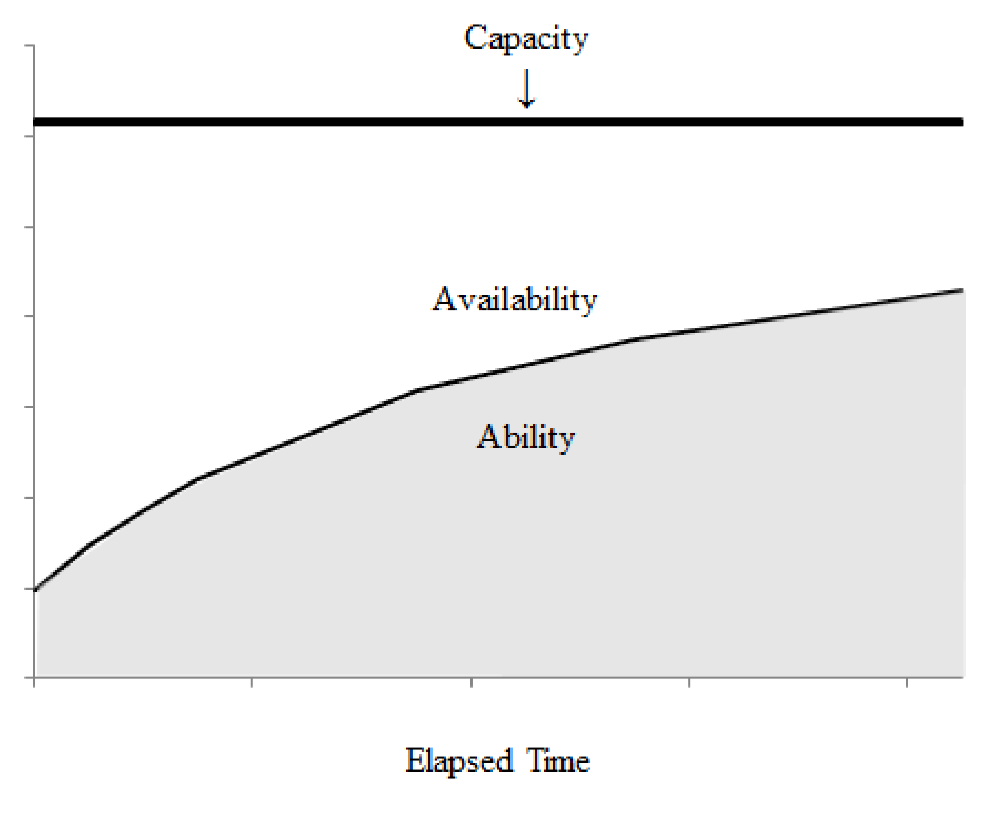
Trouble with Developing Constructs

- These previous approaches are infeasible in large-scale data-collection and analysis contexts
- For instance, DA was highly effective, but it requires extensive one-one-one interventions.
  - This is quite time and resource intensive: sample size in recent Sternberg study was only about 130
- In many real-life educational applications, there are typically several thousand examinees
  - Often, analysts may not have access to the raw data or the examinees themselves either
- Could a growth model be applied to the problem of measuring developing constructs

Potential

- One educationally relevant developing construct is *academic potential*
- Three components of potential have been identified:
  - Ability
  - Capacity
  - Availability
- In this study, we show that a non-linear growth model may be used to measure academic potential

Background on Potential

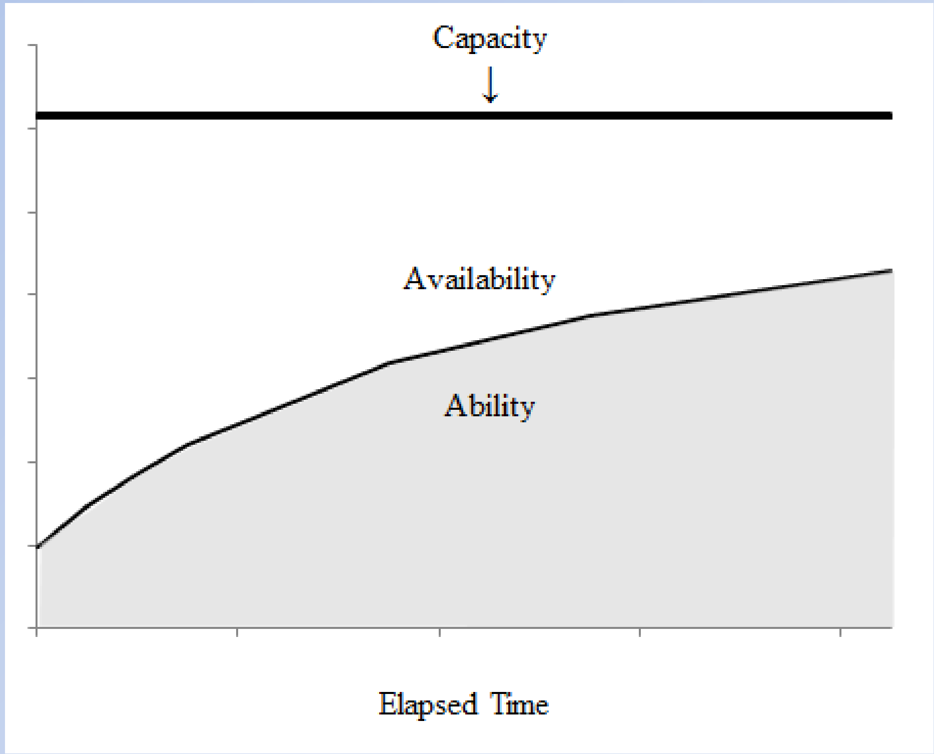


- Ability – Realized potential at time *t*
- Availability – Untapped potential at time *t*
- Capacity – The theoretical asymptote of ability as $t \rightarrow \infty$

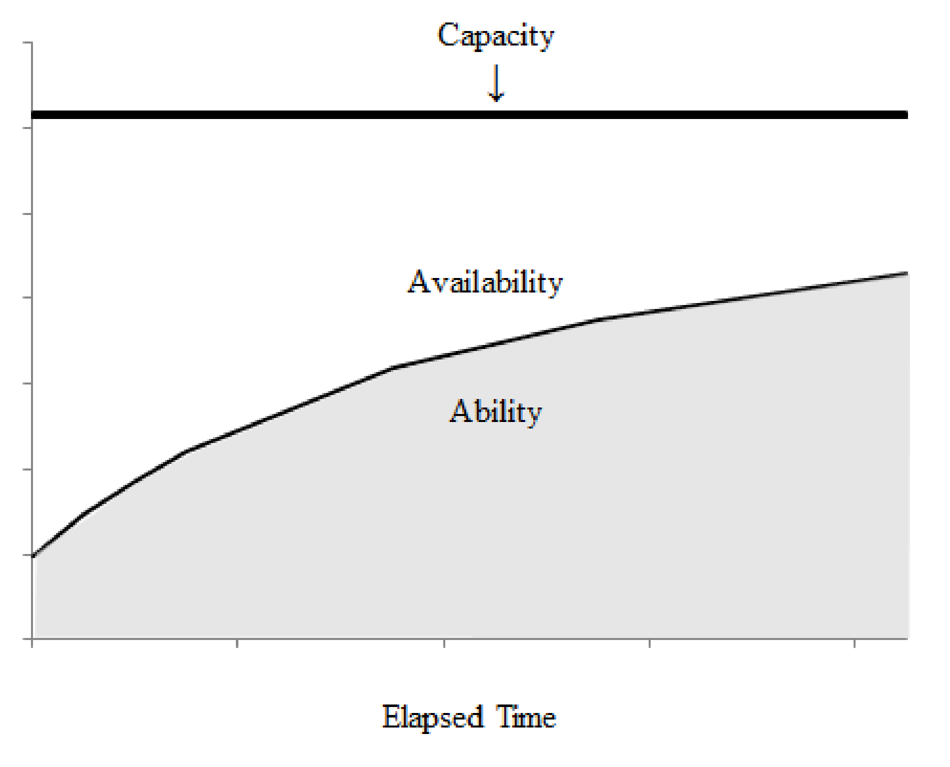
Sports Analogy

"That player has potential"

- They have a lot of room to grow (high availability)
- They have a very high ceiling (high capacity)
- They can contribute to the team now (high ability)



Background on Potential



- A non-linear mixed effect model (Gompertz, Micahelis-Menten, Exponential, etc.) may fit this type of growth:
  - Upper asymptote is capacity
  - Random effect for upper asymptote allows everyone to have unique capacity
  - Random effect on rate parameter allows everyone to learn at different rate

Using ECLS-K 1999

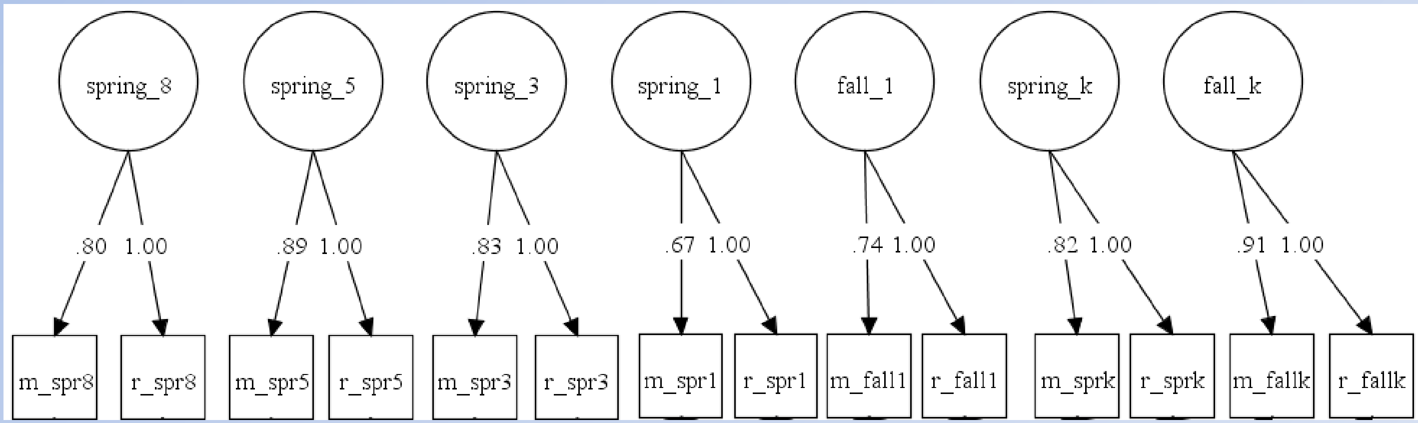
- Will walk through a simple illustrative example that models academic potential

- 1,978 kids are followed from kindergarten to 8th grade
  - 7 total time points (Fall K, Spr. K, Fall G1, Spr. G1, Spr. G3, Spr. G5, Spr. G8)

- Will demonstrate each step of the proposed model sequentially

First-Order Model

- Previous attempts have used CFA to measure potential directly
  - While the literature has gone away from CFA, we will still use CFA but as a way to measure the construct whose potential is of interest
  - Example with model "Academic Ability"
- To keep the model simple, we use vertically scaled math and reading scores at each of the 7 time points
  - Two indicators per factor



RMSEA = 0.055
CFI = 0.989
SRMR=0.045

Standardized loadings in the mid 0.80s

Longitudinal Invariance

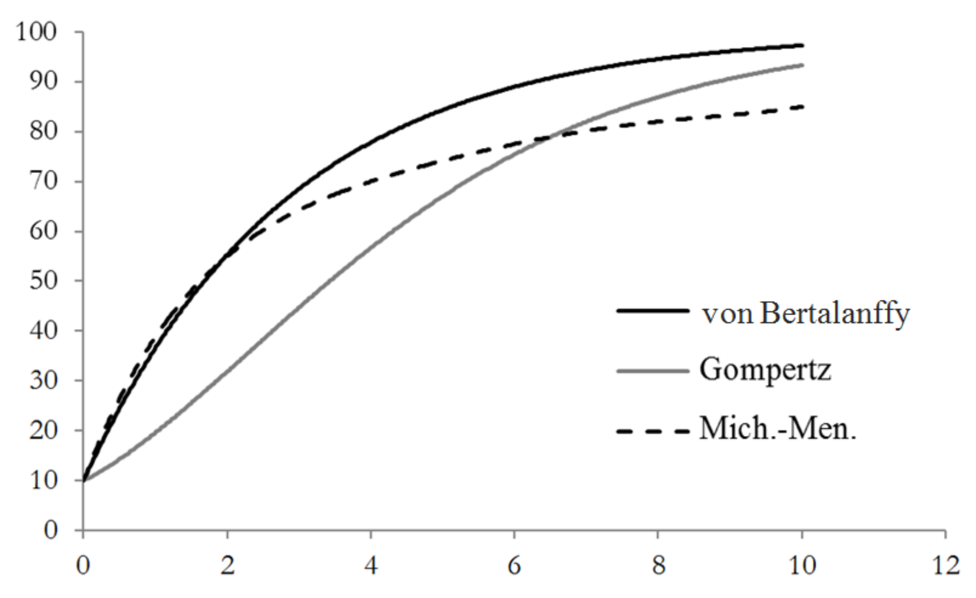
- Before moving to a growth model, loadings will be constrained to be equal across time points to test whether the latent variables are being measured the same way over time

Overall		Free	Weak	Strong
	RMSEA	0.055	0.061	0.066
	CFI	0.989	0.985	0.981

Difference		Cut-off	Free vs. Weak	Weak vs. Strong
	ΔRMSEA	< .01	0.006	0.005
	ΔCFI	> -.01	-0.004	-0.004

Second-Order Growth Model

- The second-order growth model will model the non-linear change in the first-order factors
- Using the estimated factors from the first order model, we fit population-averaged versions of Michaelis-Menten, Gompertz, and von Bertalanffy curves



Michael-Menten Model

$$Academic\ Ability_{it} = \beta_{0i} + \frac{(\beta_{Fi} - \beta_{0i})t}{\beta_{Ri} + t} + d_{it}$$

$$\beta_{0i} = \alpha_0 + \zeta_{0i}$$

$$\beta_{Fi} = \alpha_F + \zeta_{Fi}$$

$$\beta_{Ri} = \alpha_R + \zeta_{Ri}$$

$$\boldsymbol{\zeta}_i \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} g_{00} & g_{0F} & g_{0R} \\ g_{F0} & g_{FF} & g_{FR} \\ g_{R0} & g_{RF} & g_{RR} \end{bmatrix} \right)$$

$$\mathbf{d}_i \sim MVN\left(\mathbf{0}_7, diag\left(\sigma_1^{\ 2} \quad \sigma_2^{\ 2} \quad \sigma_3^{\ 2} \quad \sigma_4^{\ 2} \quad \sigma_5^{\ 2} \quad \sigma_6^{\ 2} \quad \sigma_7^{\ 2}\right)\right)$$

MM has 3 parameters

1) **Initial Value**
2) **Upper Asymptote**
3) **Mid-Point**

Ed. Psych/Intelligence researchers haven't considered reparameterization to get the capacity estimate they covet

Non-Linear Mixed Model vs. Structured Latent Curve Model

- There are two possible ways to estimate this model:
  - 1) Non-linear mixed effects model (e.g., SAS Proc NLMIXED)
  - 2) Structured latent curve model (e.g., M*plus*)

- Because we are specifically interested in subject-specific random effects, the NLME model is the only framework capable of accurately estimating these quantities (Blozis & Harring, 2016)

- The linearization that is used in the SLC model distorts the meaning of the subject-specific random effect estimates.

Model Estimates

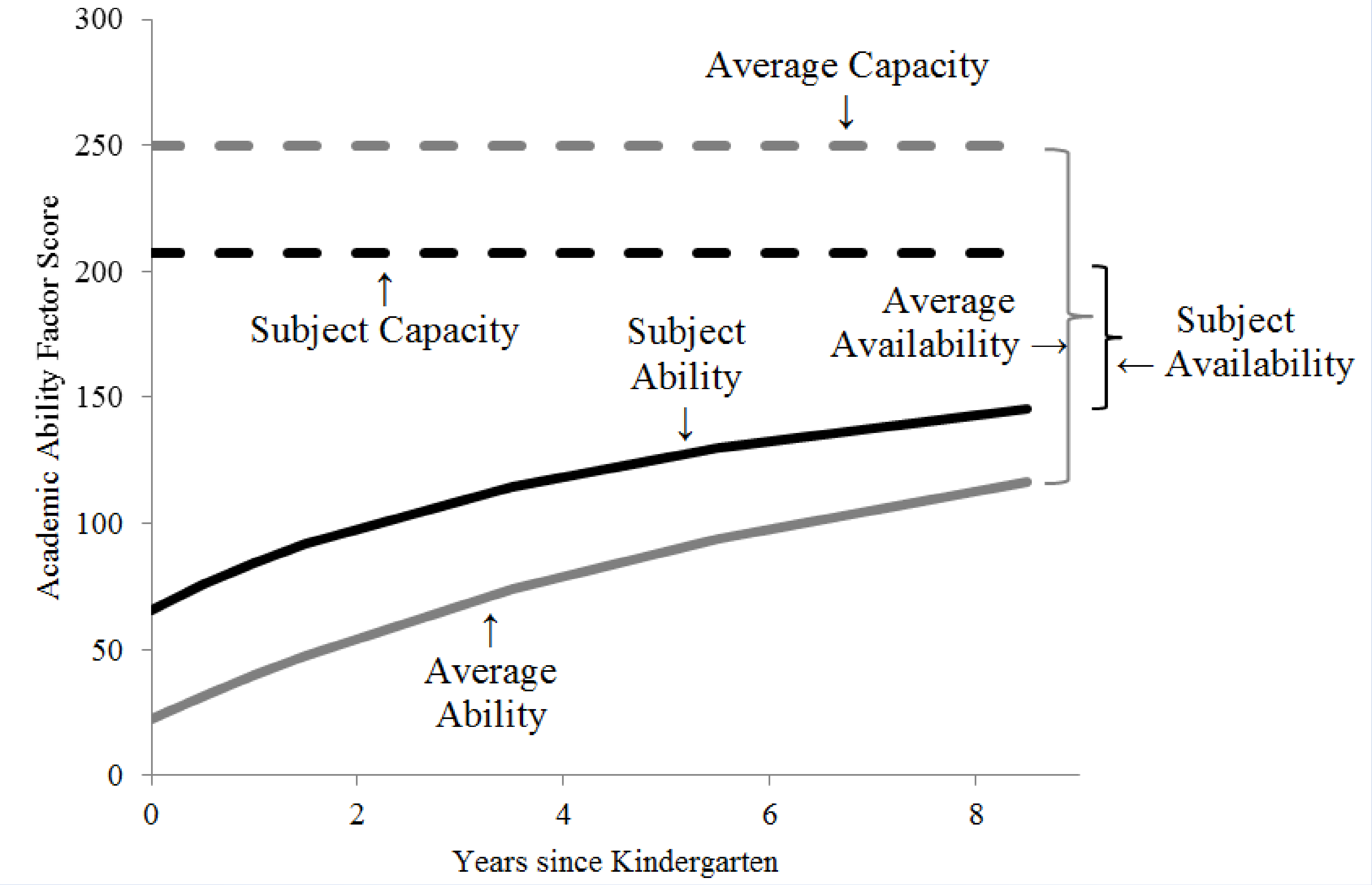
Measurement Model		
Time	Indicator	Std. Loading
Fall Kindergarten	Math	0.87
	Read	0.84
Spring Kindergarten	Math	0.85
	Read	0.82
Fall Grade 1	Math	0.83
	Read	0.79
Spring Grade 1	Math	0.84
	Read	0.79
Spring Grade 3	Math	0.88
	Read	0.85
Spring Grade 5	Math	0.87
	Read	0.86
Spring Grade 8	Math	0.90
	Read	0.83

Second-Order Growth Model	
Parameter Name	Est.
Initial Value	22.82
Capacity	249.47
Rate	9.07
Var (Initial Value)	45.12
Var (Capacity)	2498.02
Var (Rate)	9.26
Corr (Initial, Capacity)	0.24
Corr (Initial Value, Rate)	-0.15
Corr (Capacity, Rate)	0.17

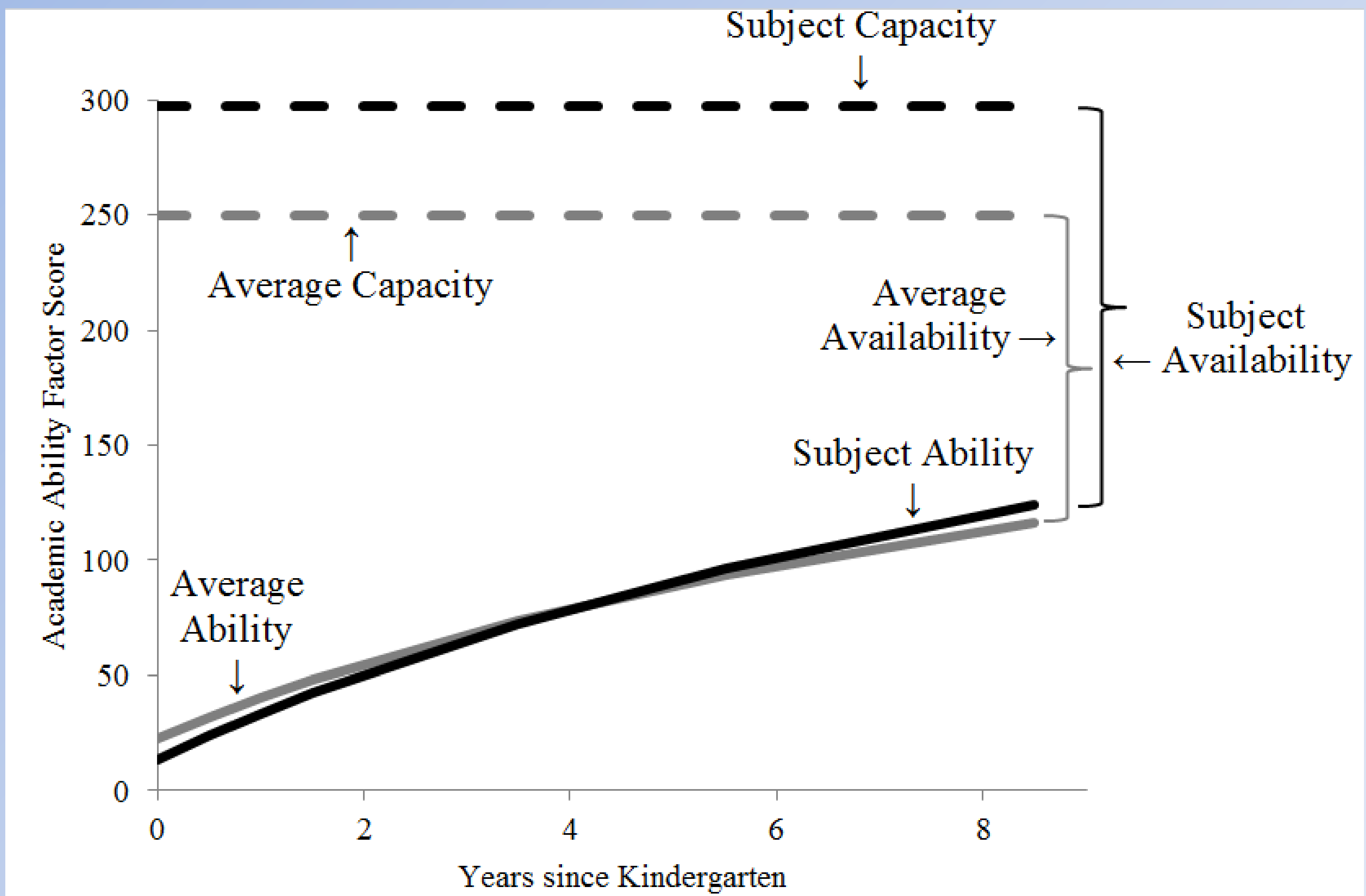
Measuring Potential

The real interest of the model is in subject-specific information to discern who has high/low capacity/availability

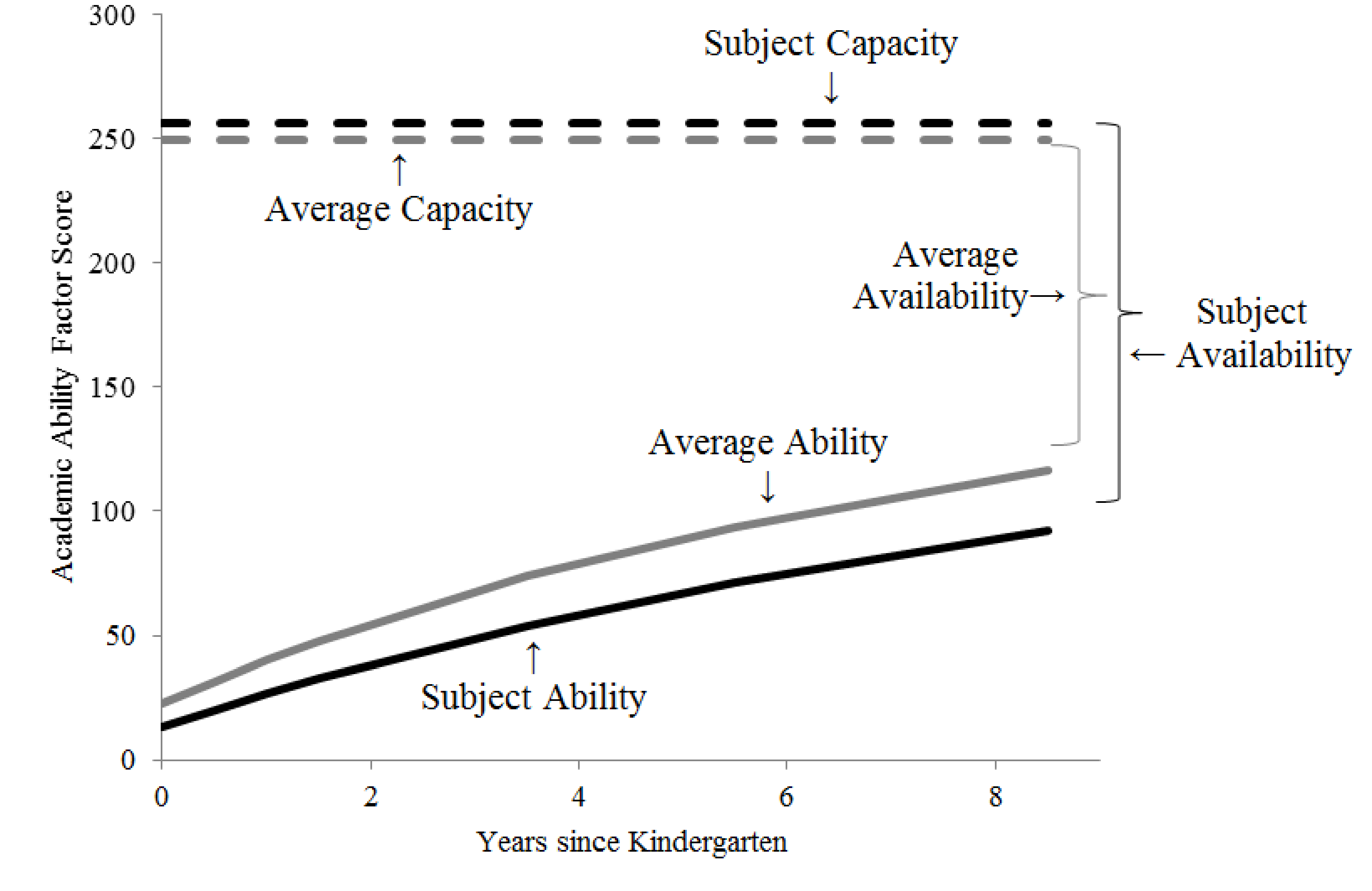
Typical “Early Bloomer”



Typical “Fast Learner”



Typical "Late Bloomer"



Construct Validity

- A common criticism of educational assessments is that they unfairly advantage certain demographic groups
  - Hurts construct validity because measures are unduly effected by variables like SES
- We compare the results of our proposed model by conducting general linear models with Sex, Race, SES, and all possible two and three way interactions.
  - SES was measured by conducting a PCA on many SES-related variables
- We then compare these to the results from the developed constructs in the ECLS-K data

GLM Results

Capacity Scores			
Source	F	p	ω^2
Sex	0.09	.77	.00
Ethnicity	2.10	.01	<.01
SES	10.59	<.01	<.01
Sex×Eth	0.41	.74	.00
SES×Sex	0.05	.82	.00
SES×Eth	1.69	.17	.00
SES×Sex×Eth	0.26	.86	.00
R^2			.033

8th Grade Reading Scores			
Source	F	p	ω^2
Sex	1.62	.20	.00
Ethnicity	9.65	<.01	.01
SES	125.71	<.01	.05
Sex×Eth	1.19	.31	.00
SES×Sex	2.12	.15	.00
SES×Eth	0.73	.53	.00
SES×Sex×Eth	1.32	.27	.00
R^2			.240

GLM Results

The effect size for SES on Capacity scores is 90% lower than the effect of SES on reading scores in 8th grade.

The same pattern holds for math scores in 8th grade as well (not shown)

	Capa				ding Scores		
Source	*F*					*p*	ω^2
Sex	0.09					.20	.00
Ethnicity	2.10					<.01	.01
SES	10.59					<.01	.05
Sex×Eth	0.41					.31	.00
SES×Sex	0.05					.15	.00
SES×Eth	1.69	.17	.00	SES×Eth	0.73	.53	.00
SES×Sex×Eth	0.26	.86	.00	SES×Sex×Eth	1.32	.27	.00
R^2			.033	R^2			.240

GLM Results

Kindergarten Availability			
Source	*F*	*p*	ω^2
Sex	0.38	.53	.00
Ethnicity	1.99	.11	<.01
SES	1.19	.27	.00
Sex×Eth	0.50	.68	.00
SES×Sex	0.07	.79	.00
SES×Eth	1.71	.16	<.01
SES×Sex×Eth	0.39	.76	.00
R^2			.018

Kindergarten Reading			
Source	*F*	*p*	ω^2
Sex	15.11	.00	.01
Ethnicity	6.22	.00	.01
SES	158.09	<.01	.07
Sex×Eth	3.19	.02	.00
SES×Sex	0.11	.74	.00
SES×Eth	6.45	.00	.01
SES×Sex×Eth	0.53	.66	.00
R^2			.167

Universiteit Utrecht

GLM Results

	Kindergarten A…	…ading
Source	F	ω^2
Sex	0.38	.01
Ethnicity	1.99	.01
SES	1.19	.07
Sex×Eth	0.50	.00
SES×Sex	0.07	.00
SES×Eth	1.71	.01
SES×Sex× Eth	0.39	.00
R^2		.167

The effect size for SES on Availability scores is at Kindergarten is 90% lower than the effect of SES on reading scores in kindergarten.

Even with the large sample, no statistical significance for availability scores.

The same pattern holds for math scores in Kindergarten as well (not shown)

Discussion and Implications

- Previous attempts to quantify potential have used pre/post-test interventions or short intervals which do not give the whole picture
  - Extrapolating from developed constructs is also a suboptimal strategy
- The Capacity/Ability/Availability framework introduced here shows that all three components of potential can be incorporated simultaneously
- Although CFA is not sufficient for developing constructs, adding a second-order growth model can vastly simplify measurement
  - The addition of the second order growth model extends the measurement framework from developed to developing constructs
- Could also use the Capacity or Availability estimates as an outcome or predictor in other models
  - Could be highly useful in some areas as a control variable (e.g., teacher effectiveness models)

Limitations and Caveats

- Given the historic relationship between psychometrics and the eugenics movement, ethical considerations must always be front-and-center
- This model is NOT tapping into genetic dispositions
- Capacity itself is also likely malleable depending on educational opportunity
- Capacity estimates are still based on test scores
  - Measures of potential are therefore limited to what the tests themselves are assessing
- Intended use is to help identify students who could most benefit from access to additional resources and/or interventions
  - Who is not living up to their potential and how can we help them?
- Although our example was admittedly tame, this method shows promise for the large-scale measurement of potential.

Limitations and Caveats

- Must model the trajectory of the second-order growth correctly
- The second-order model will not be effective if the first-order model has poor measurement quality
  - Poor measurement quality could also introduce convergence issues
- The further away the observed time is from the asymptotic behavior, the larger the uncertainty of capacity will be
- Nothing yet about predictive validity – e.g., does capacity better predict college performance than ACT/SAT?