Measuring financial literacy with the big three

Annamaria Lusardi (George Washington University)
Tim Kaiser (University of Kaiserslautern-Landau)
Olivia S. Mitchell (University of Pennsylvania)
Luis Oberrauch (University of Kaiserslautern-Landau)

G53 Network Conference
June, 2023

Main question and relevance

- Measuring financial literacy via three (or five) financial knowledge test items has become the norm in a large body of work in applied microeconomics (see Lusardi and Mitchell 2014, JEL for an overview).

- Despite their popularity, the use of these items is sometimes criticized by scholars from adjacent disciplines (education, psychology, etc.)

  → The main criticism stems from the lack of psychometric evidence regarding item characteristics.

  → How valid are the big three (five) financial literacy items?

This paper

Psychometric evaluation of the most used financial knowledge items using data from the 2018 National Financial Capability Study (RAND-ALP)

- Psychometric criteria suggest adequate scale validity irrespective of number of items used (three vs. five)
  - No evidence of item bias „Differential Item Functioning“ (DIF)
  - High discrimination between low and high achievers
  - Unidimensionality

- FL measure with predictive validity regarding financial behaviors
- FL correlates with respondent-level variables known from previous literature (criterion validity)

Previous literature

Narrative review of measurement scales (Huston 2010, p.309)

"Thus, initial instruments consisting of as few as three items (Henry, Weber, and Yarbrough 2001; Lusardi 2008a; Lusardi and Mitchell 2007a, 2007c, 2008c) would appear to be deficient to capture the breadth of human capital specifically related to personal finance. After initial testing, techniques such as item response theory approaches could be used to reduce the number of items." (Edelen et al. 2006)

Previous literature

Developement and psychometric validation of more extensive scales

- Knoll and Houts (2012, JCA): 20 items (IRT)
- Knoll and Houts (2020, JCA): 10 items (IRT)
- Fernandes et al. (2014, ManSci): 13 items (CTT)

Is there a need for more extensive item sets?

“

Data

	N	Mean	SD	Min	Max
Demographics					
Male	1233	0.418		0	1
Age	1233	55.887	14.709	22	96
Non-white	1233	0.175		0	1
Married	1233	0.588		0	1
Widowed	1233	0.069		0	1
Children	1214	0.353		0	1
No college education	1233	0.159		0	1
Risk seeking	1195	4.635	2.425	1	10
Job status					
Employed	1233	0.562		0	1
Unemployed	1233	0.041		0	1
Retired	1233	0.317		0	1
Homemaker	1233	0.087		0	1
Income					
< 25,000$	1231	0.075		0	1
25,000$ ≤ income < 50,000$	1231	0.326		0	1
50,000$ ≤ income < 75,000$	1231	0.224		0	1
≥ 75,000$	1231	0.375		0	1
Recent income drop	1210	0.15		0	1

Rand ALP (National Financial Capability Study, 2018, Wave 2)

Financial literacy

Item	N	% correct
Numeracy Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow? a) More than $102 b) Exactly $102 c) Less than $102	1207	85.3 %
Inflation Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account? a) More than today b) Exactly the same c) Less than today	1207	75.6 %
Bonds If interest rates rise, what will typically happen to bond prices? a) They will rise b) They will fall c) They will stay the same d) There is no relationship between bond prices and the interest rate	1208	34.9 %
Mortgage A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less. a) True b) False	1208	83.9 %
Risk diversification Buying a single company's stock usually provides a safer return than a stock mutual fund a) True b) False	1207	62.2 %

IRT Appraoch

$$P(\mathrm{X}_j = 1 \mid \theta_v, \sigma_i, \alpha_i, \gamma_i, \delta_i) = \gamma_i + (\delta_i - \gamma_i)\ \frac{\exp[\alpha_i(\theta_v - \sigma_i)]}{1 + \exp[\alpha_i(\theta_v - \sigma_i)]}$$

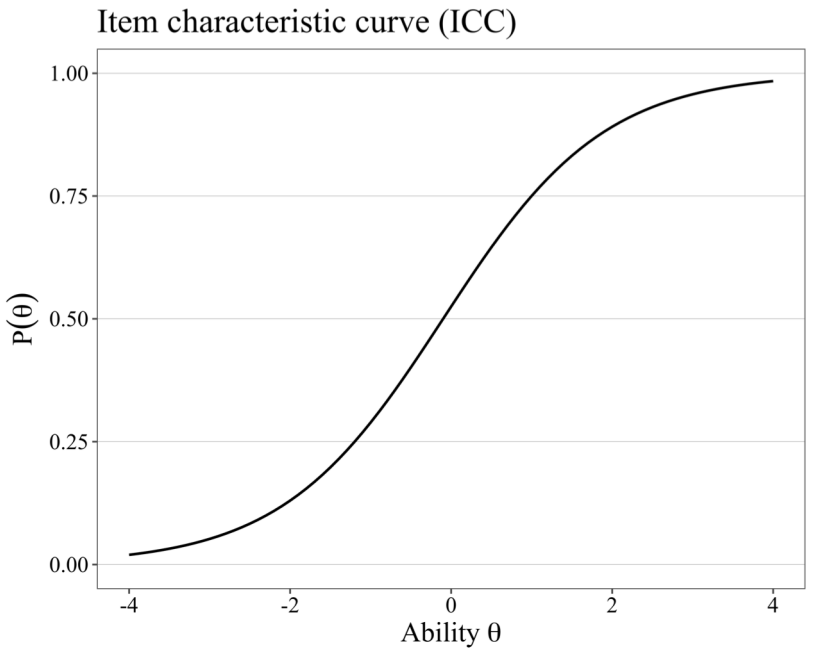
θ_v : Person‘s ability

σ_i : Item difficulty (Location)

α_i : Item discrimination (slope)

γ_i : Guessing parameter (lower asymptote)

δ_i : Inattention parameter (upper asymptote)



IRT Appraoch

$$P(\mathrm{X}_j = 1 \mid \theta_v, \sigma_i, \alpha_i, \gamma_i, \delta_i) = \gamma_i + (\delta_i - \gamma_i)\ \frac{\exp[\alpha_i(\theta_v - \sigma_i)]}{1 + \exp[\alpha_i(\theta_v - \sigma_i)]}$$

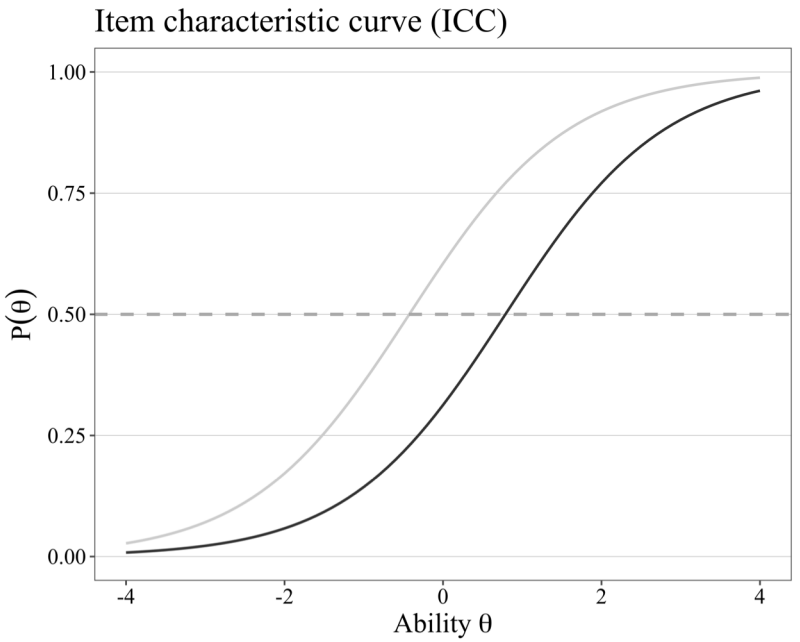
θ_v : Person‘s ability

σ_i : Item difficulty (Location)

α_i : Item discrimination (slope)

γ_i : Guessing parameter (lower asymptote)

δ_i : Inattention parameter (upper asymptote)



IRT Appraoch

$$P(\mathrm{X}_j = 1 \mid \theta_v, \sigma_i, \alpha_i, \gamma_i, \delta_i) = \gamma_i + (\delta_i - \gamma_i)\ \frac{\exp[\alpha_i(\theta_v - \sigma_i)]}{1 + \exp[\alpha_i(\theta_v - \sigma_i)]}$$

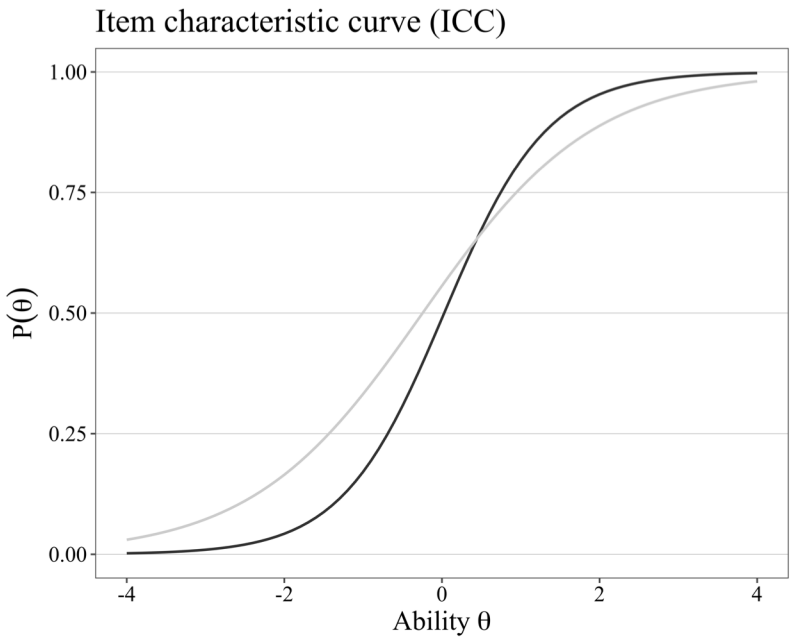
θ_v : Person‘s ability

σ_i : Item difficulty (Location)

α_i : Item discrimination (slope)

γ_i : Guessing parameter (lower asymptote)

δ_i : Inattention parameter (upper asymptote)



IRT Appraoch

$$P(\mathrm{X}_j = 1 \mid \theta_v, \sigma_i, \alpha_i, \gamma_i, \delta_i) = \gamma_i + (\delta_i - \gamma_i)\ \frac{\exp[\alpha_i(\theta_v - \sigma_i)]}{1 + \exp[\alpha_i(\theta_v - \sigma_i)]}$$

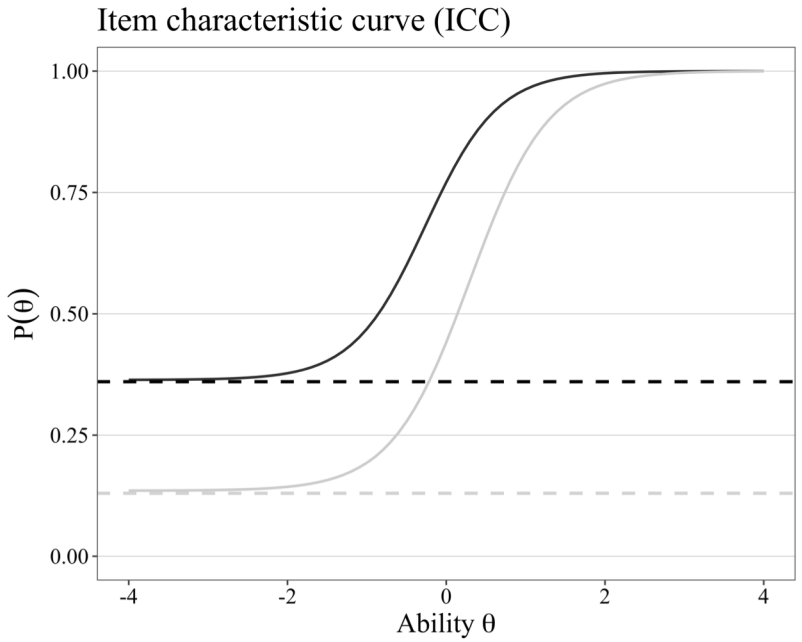
θ_v : Person‘s ability

σ_i : Item difficulty (Location)

α_i : Item discrimination (slope)

γ_i : Guessing parameter (lower asymptote)

δ_i : Inattention parameter (upper asymptote)



IRT Appraoch

$$P(\mathrm{X}_j = 1 \mid \theta_v, \sigma_i, \alpha_i, \gamma_i, \delta_i) = \gamma_i + (\delta_i - \gamma_i)\ \frac{\exp[\alpha_i(\theta_v - \sigma_i)]}{1 + \exp[\alpha_i(\theta_v - \sigma_i)]}$$

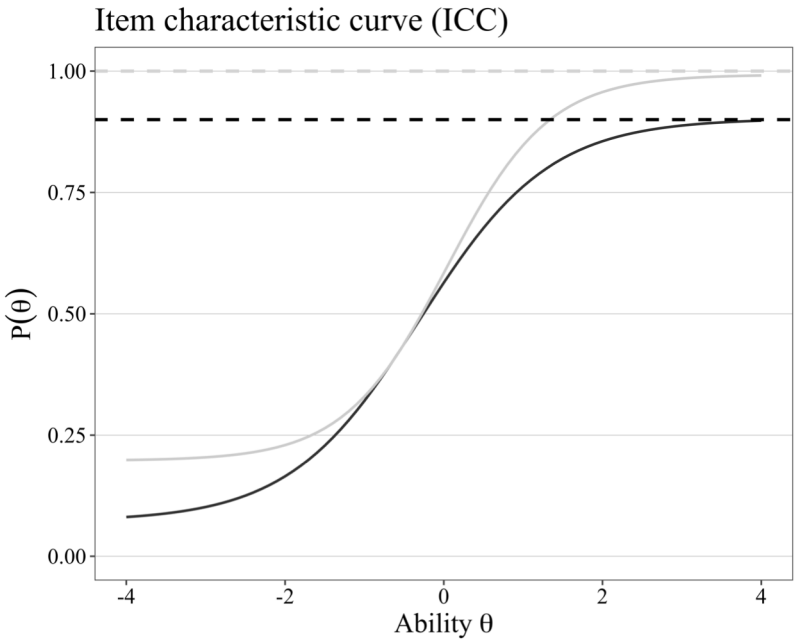
θ_v : Person‘s ability

σ_i : Item difficulty (Location)

α_i : Item discrimination (slope)

γ_i : Guessing parameter (lower asymptote)

δ_i : Inattention parameter (upper asymptote)



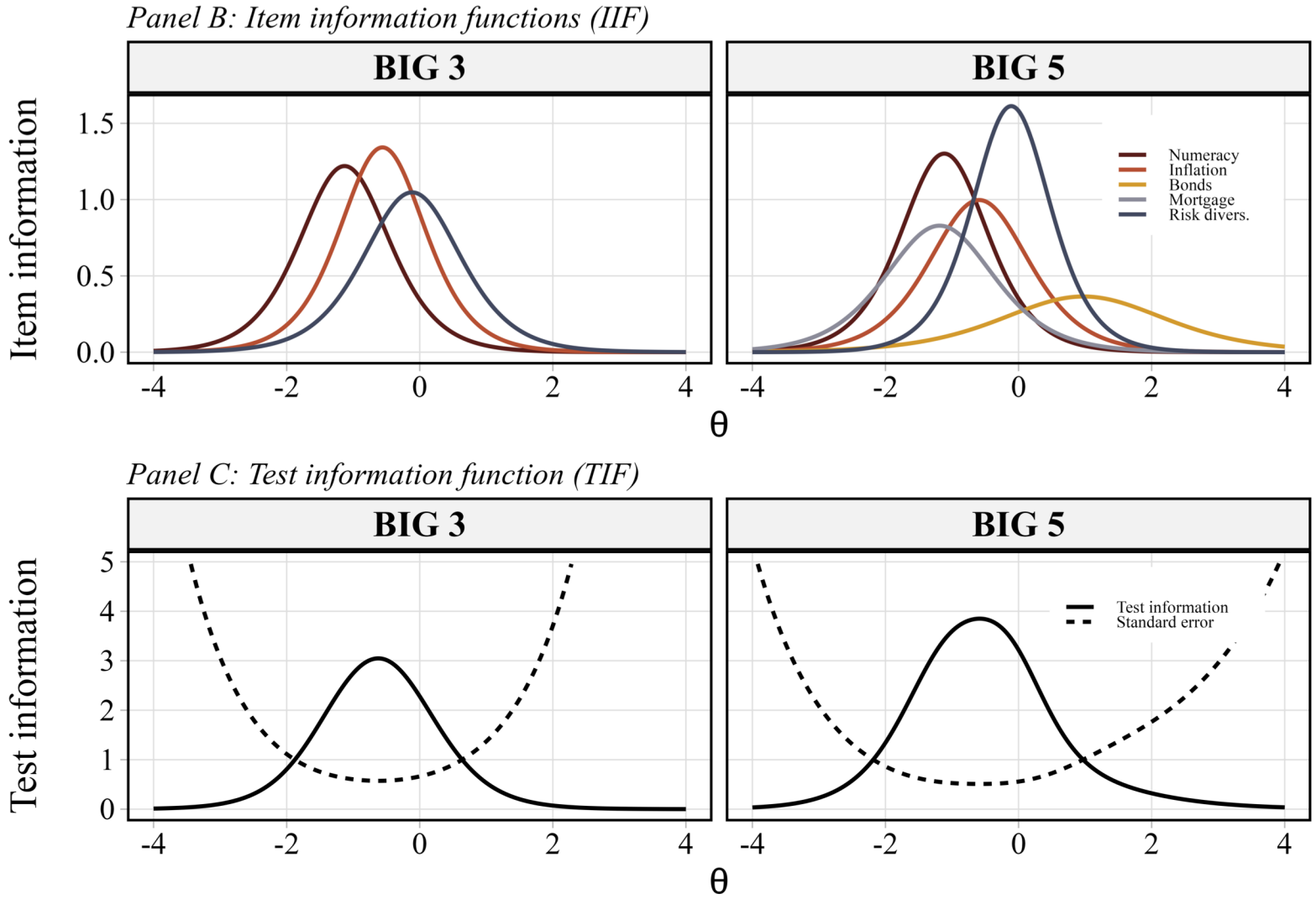
Results: How many parameters?

	1-PL *(SRMSR=0.043)*		2-PL *(SRMSR=0.034)*		3-PL *(SRMSR=0.033)*		4-PL *(SRMSR=0.031)*	
Item	$S\text{-}\chi^2$	p-val.	$S\text{-}\chi^2$	p-val.	$S\text{-}\chi^2$	p-val.	$S\text{-}\chi^2$	p-val.
Numeracy	0.963	0.810	2.991	0.224	7.892	0.005	6.989	0
Inflation	4.340	0.227	5.669	0.059	4.260	0.039	4.473	0
Bonds	25.294	0.000	3.192	0.203	3.793	0.051	4.217	0
Mortgage	6.154	0.104	4.582	0.101	7.500	0.006	6.684	0
Risk divers.	17.273	0.001	12.050	0.002	0.896	0.344	0.948	0

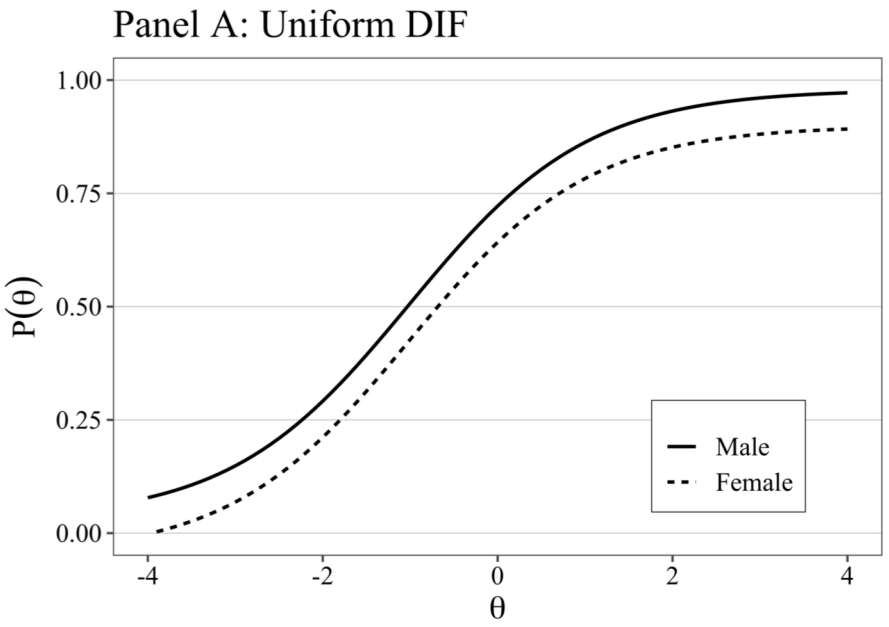
Results: Item characteristics

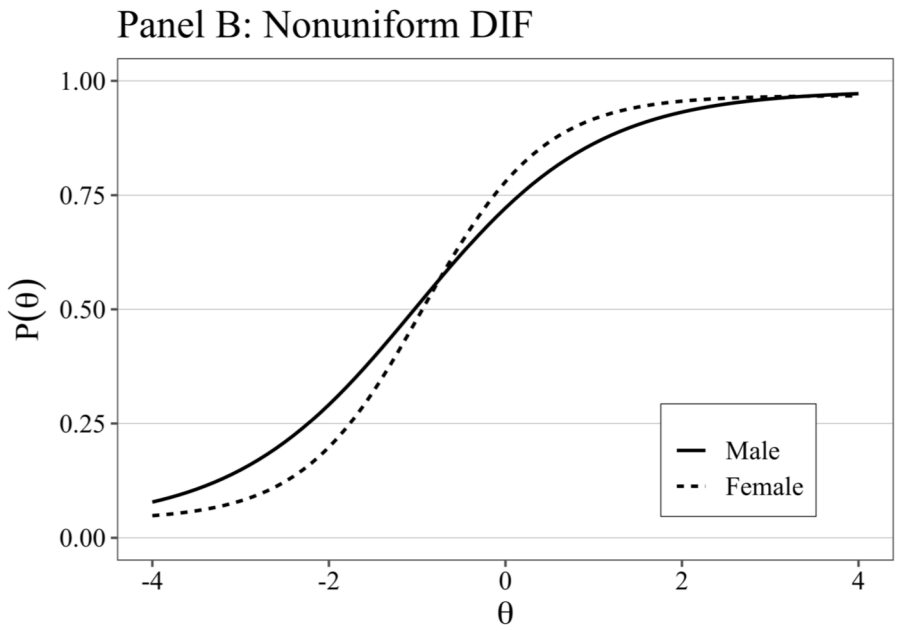
		CTT		IRT (BIG 3)		IRT (BIG 5)	
Item	n	Freq.	ITC	$\hat{\alpha}$ [SE]	$\hat{\alpha}$ [SE]	a [SE]	$\hat{\sigma}$ [SE]
Numeracy	1208	0.853	0.345	2.208 [0.125]	-1.13 [0.042]	2.281 [0.13]	-1.116 [0.041]
Inflation	1207	0.758	0.452	2.318 [0.133]	-0.56 [0.035]	1.997 [0.115]	-0.592 [0.039]
Bonds	1208	0.349	0.375			1.208 [0.071]	0.974 [0.059]
Mortgage	1207	0.841	0.399			1.821 [0.1]	-1.187 [0.048]
Risk divers.	1207	0.622	0.512	2.046 [0.119]	-0.115 [0.037]	2.54 [0.146]	-0.11 [0.032]

Results: Item and test information functions



Results: Differentional item functioning (DIF)





Classification of DIF-effects using the well-established schemes (ETS; Jodoin & Gierl 2001)

Category A: Negligible DIF

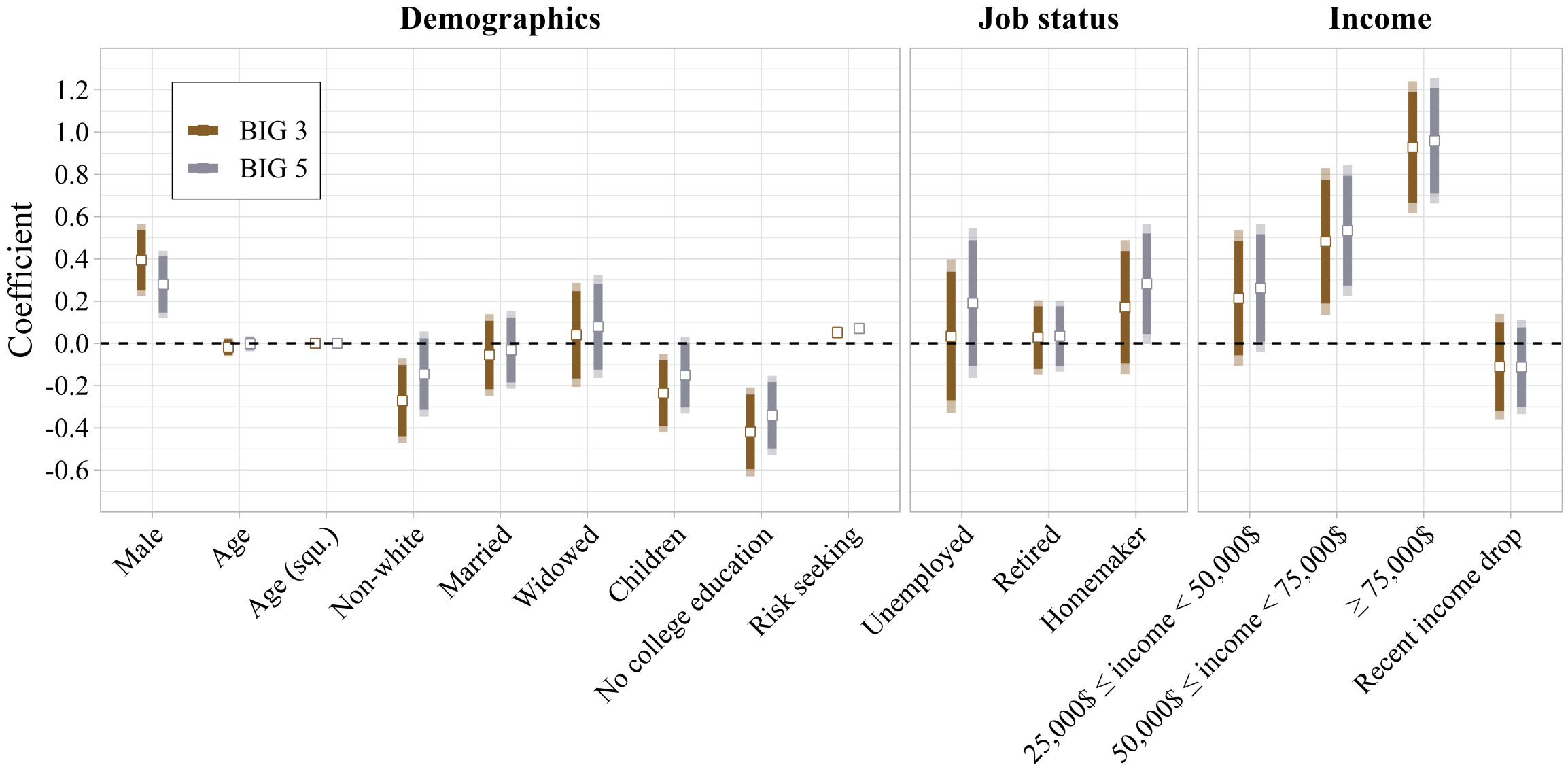
Category B: Moderate DIF

Category C: Severe DIF

Results: Differentional item functioning (DIF)

	(1)		(2)		(3)		(4)		(5)	
	Gender *Focal group: female*		Employment *Focal group: unemployed*		Education *Focal group: no college*		Race *Focal group: black*		Income *Focal group: income < 50,000$*	
	ΔR^2	J-G	ΔR^2	J-G	ΔR^2	J-G	ΔR^2	J-G	ΔR^2	J-G
Panel A: Uniform DIF										
Numeracy	0.0003	A	0.0097	A	0.0008	A	0.0031	A	0.0001	A
Inflation	0.0019	A	0.0004	A	0.0008	A	0.0004	A	0.0003	A
Bonds	0.0073	A	0.0005	A	0.007	A	0.0052	A	0.0075	A
Mortgage	0.0002	A	0.0012	A	0.0007	A	0.002	A	0.0018	A
Risk divers.	0.003	A	0	A	0.0011	A	0.0022	A	0.0009	A
Panel B: Non-uniform DIF										
Numeracy	0.0003	A	0.0024	A	0.0005	A	0	A	0	A
Inflation	0.0011	A	0.0003	A	0.0006	A	0.0004	A	0.0001	A
Bonds	0.0018	A	0.0002	A	0.0019	A	0.001	A	0.0002	A
Mortgage	0.0002	A	0	A	0.0007	A	0.0004	A	0.0006	A
Risk divers.	0.0001	A	0	A	0.0011	A	0	A	0	A

Results: Correlates of FL



Results: Predictive validity (1/2)

	(1) Retirement saving	(2) Financial satisfaction	(3) Emergency fund	(4) Financial confidence	(5) Credit score
Panel A: Cross-sectional estimates using FL scores from 2018					
(i): Big 3					
FL BIG 3	0.055 * [0.03]	0.164 [0.158]	0.077** [0.03]	0.118*** [0.038]	0.318*** [0.076]
N	832	1159	1165	1166	1136
Adj.R-Squ.	0.202	0.272	0.229	0.313	0.346
F-stat.	10.812	18.645	21.09	21.309	19.971
(ii) Big 5					
FL BIG 5	0.052** [0.024]	0.176 [0.107]	0.055** [0.023]	0.093*** [0.022]	0.214*** [0.053]
N	832	1159	1165	1166	1136
Adj.R-Squ.	0.205	0.274	0.228	0.317	0.342
F-stat.	11.032	18.945	21.778	22.504	20.001

Results: Predictive validity (2/2)

	(1) Retirement saving	(2) Financial satisfaction	(3) Emergency fund	(4) Financial confidence	(5) Credit score
Panel B: Predictive validity using FL scores from 2012					
(i) Big 3					
FL BIG 3	0.055 * [0.032]	0.169 [0.171]	0.074** [0.029]	0.067 * [0.036]	0.167** [0.079]
N	830	1157	1163	1163	1130
Adj.R-Squ.	0.204	0.27	0.227	0.293	0.337
F-stat.	10.696	18.972	17.74	18.79	21.892
(ii) Big 5					
FL BIG 5	0.058 * [0.03]	0.092 [0.139]	0.062** [0.025]	0.06** [0.028]	0.147** [0.067]
N	830	1157	1163	1163	1130
Adj.R-Squ.	0.206	0.269	0.226	0.293	0.337
F-stat.	10.765	18.861	17.4	19.403	21.877

Summary

Measuring financial literacy via the ”big three” or “big five” financial knowledge test items is backed by psychometric evidence

- Relying on the 3-item scale works well, albeit with some loss of precision
- Limited test information for high ability respondents when using the 3-item scale
- Concurrent and predictive validity confirmed for both item sets (with larger standard errors when relying on only 3 items)

→**The big three (five) financial literacy items work well, especially when it comes to identifying individuals with below-average financial literacy.**

Appendix

DIF effects based on χ^2-statistics

	Gender		Employment		Education		Race	
Item	Lord [MH]	ETS	Lord [MH]	ETS	Lord [MH]	ETS	Lord [MH]	ETS
Numeracy	-0.1455 [0.0619]	A	0.9628 [-0.4097]	A	0.1939 [-0.0825]	A	0.1314 [-0.0559]	A
Inflation	-0.4491 [0.1911]	A	-0.7687 [0.3271]	A	-0.4801 [0.2043]	A	0.3910 [-0.1664]	A
Bonds	0.7081 [-0.3013]	A	0.5826 [-0.2479]	A	0.7950 [-0.3383]	A	-0.6799 [0.2893]	A
Mortgage	0.2482 [-0.1056]	A	-0.0754 [0.0321]	A	0.3476 [-0.1479]	A	-0.2554 [0.1087]	A
Risk divers	-0.3619 [0.1540]	A	-0.7012 [0.2984]	A	-0.8563 [0.3644]	A	0.4129 [-0.1757]	A

Testing against alternative models

Testing against two-dimensional model

Model	loglike	Deviance	Nobs	AIC	BIC	GHP
Missing response model	-2,526.0	5,052.0	1,233	5,084.0	5,165.8	0.7
2-PL IRT model	-1,955.8	3,911.6	1,233	3,923.6	3,954.3	0.5

Testing against partial credit model (PCM)

Model	loglike	Deviance	Nobs	AIC	BIC	GHP
Partial crecit model (PCM)	-4341.86	8683.718	1208	8723.718	8825.6	0,72
2-pl IRT model	-3074.59	6149.188	1208	6169.188	6220,1	0,51