Inequality in noncognitive skills: a cross-cohort comparison EXTENDED ABSTRACT

Orazio Attanasio[1], Richard Blundell[1], Gabriella Conti[1], and Giacomo Mason[1]

[1]*UCL Department of Economics, London, UK*

November 20, 2017

In this study, we derive measures of noncognitive ability at age 5 for two large cohorts of British children that are comparable across time. We use these measures to document that socioeconomic inequality in noncognitive ability has substantially increased between 1975 and 2005.

In the vast economic literature on the returns to human capital, the original focus on cognitive ability has recently shifted to encompass personality traits and social/behavioural skills – sometimes defined collectively as "noncognitive ability" (Cunha and Heckman, 2007; Cunha et al., 2010). There is now substantial agreement that such "soft" skills earlier in life play a prominent role in shaping adult outcomes such as education, occupation, and earnings (Heckman et al., 2006; Carneiro et al., 2007; Heckman and Kautz, 2012). Furthermore, noncognitive skills seem to be malleable until children are older, and have been shown to be perhaps the most important channel through which early childhood development interventions lead to long-term results (Heckman et al., 2013; Kautz et al., 2014).

Socioeconomic gradients in behaviour skills have been documented, and are persistent across childhood and adolescence (Campbell, 1995; Segal, 2008). However, it is not clear how and whether inequalities in noncognitive ability are stable or change in time. The usual approach is to just assume that measures collected from different populations at different points in time using different instruments are directly comparable using simple points-based scores. This work utilises data from two of the main British cohorts to shed light on the evolution of socioeconomic gradients in the UK through time, with a particular focus on the rigorous comparison of skills as measured by maternal reports on children's behaviours. Cross-cohort comparisons have been performed in the past. Blanden et al. (2007) finds that noncognitive ability can explain up to 20% of income persistence across the 1958 and 1970 birth cohorts; however, no attempt was made to assess whether the construct in the two cohorts were measured in a comparable way.

We use data from two cohorts of children from the United Kingdom. The first is the British Cohort Study (BCS), which recruited more than 17,000 people born across the UK in a single week in 1970. The second source of data is the Millennium Cohort Study (MCS), which follows more than 19,000 children born in the UK in 2000-01. From these studies, we select the subsample of children who were born in England. We account for the complex, stratified survey design of MCS by "rebalancing" the sample: we randomly drop observations from each stratum according to the sampling probabilities reported in Plewis (2007). Our sample of children with nonmissing data on child behaviour has 9676 in BCS and 5710 in MCS. For both cohorts, our main focus is maternal reports of the child's behaviours collected at five years of age. We merge these reports with data on weekly total family net income when the child was of age 10; this information has been harmonised across cohorts and derived as part of CLOSER, which is funded by the ESRC

Maternal reports on child behaviours were surveyed with two different instruments in the two cohorts. For BCS, the Rutter A Scale was used (Rutter et al., 1970). The mother is presented with a list of behaviours that a child might exhibit, and is asked for each whether it 'Does not apply', it 'Somewhat applies', or it 'Certainly applies' to her child. In the MCS cohort, mothers were administered the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1994, 1997). The SDQ was created as an updated version of the Rutter scales, expanding it to more recent advances in child psychopathology Stone et al. (2010). Similarly to the Rutter scale, it asks the parent to rate the child on different behaviours, with the options of 'Not true', 'Somewhat true', 'Certainly true'. Summary scores obtained from the SDQ has been shown to correlate highly with the ones from Rutter scales, but the invariance properties of the scales are sensitive to the wording of the questions (Goodman et al., 2007).

Our aim is to assess whether the two scales can be used to derive measures of noncognitive ability that are invariant, and can therefore be compared across cohorts. In the first step of our analysis, we consider a subset of each scale's items, selecting only those that present very similar wording across scales. The selection process results in a new 12-item scale, reported in Table 1 with the original wording and the prevalence of each option. Some of the items are coded differently in the two original scales (e.g. they are positively coded in the SDQ like 6 and 7, or only binary in the Rutter scales like 5 and 12), and we thus turn them into binary dummies to ensure comparability.

The second step to our invariance analysis seeks to justify a common latent factor structure across the two cohorts. The Rutter scale is constructed to measure two latent constructs: anti-social (externalising problems), and neurotic (internalising problems) behaviours. The five-factor structure at the basis of the full SDQ has been validated in many contexts (Stone et al., 2010); however, lower-dimensional structures have been also suggested (Dickey and Blumberg, 2004), and recent research has shown that there are some benefits to using broader subscales that correspond to the externalising and internalising factors in Rutter (Goodman et al., 2010). To assess whether the two-factor conceptualisation suggested by the theory is valid in our "reduced" 12-item scale, we conduct exploratory factor analysis (EFA) on the new 12-item scale separately for each cohort sample. Different methodologies for quantifying the optimal number of factors produce inconclusive results, but both Velicer MAP and very simple structure (VSS) seem to indicate that a two-factor structure is suitable. An EFA model with two factors, employing the usual cutoff of .3 for the loadings, delivers a neat and sensible separation between items, as shown in the second column of Table 1. Thus, our two-factor structure seems justifiable on both theoretical and empirical grounds.

Once the dimensionality of the problem has been assessed across the two cohorts, we can proceed to the last step of the analysis, where we test measurement invariance. To this end, we set up a confirmatory categorical factor analysis model that incorporates the congeneric (dedicated) factor structure established above, where items 1-7 load on the EXT latent factor and items 8-12 load on INT. Measurement invariance is usually tested by starting from the least restrictive indentifying model, estimating a series of models with increasing restrictions on the model parameters across the groups (cohorts, in our case), and testing whether fit worsens relatively to the unrestricted model. We follow the approach in recent work by Wu and Estabrook (2016), who present a comprehensive identification analysis for models of different invariance levels. This methodology is slightly different from the one commonly used in the ordered categorical measurement invariance literature (Millsap and Tein, 2004); we are not aware of any empirical applications of this new approach. We find that a model that restricts item thresholds and loadings fits the data as well as the unrestricted model. This enables us to make meaningful comparisons of the latent factors on the same scale. However, invariance of the intercept parameters – which is needed to compare the means of latent factors across groups – is rejected.

The model with restricted loadings and thresholds can nevertheless be used to investigate how inequality of noncognitive skills has changed between the two cohorts. The variances of the EXT and INT factors are estimated to be larger in the MCS children by .24 and .73 of a standard deviation, respectively, indicating larger dispersion. We can then use the model to score children's level on the two latent dimensions, and examine them against the socioeconomic

status of their families. In Figure 1, we can see that the gradient in noncognitive skills along family income quintiles is more pronounced in the MCS cohort. In particular, a five-year old at the top quintile of the income distribution in 1975 has a better score by 0.25 in the externalising dimension and 0.15 in the internalising dimension, compared to a child at the bottom quintile. In 2005, the differences amount to 0.53 and 0.44, respectively.

To our knowledge, our study is the first to use measurement invariance analysis to assess the comparability of maternal reports on child behaviours across different UK cohorts. We find that a subset of items from two different instruments can be used to compare children on two latent dimensions of externalising and internalising problems. The measures we obtain show a marked increase in inequality between 1975 and 2005. These measures can be used in future research to estimate how the returns to noncognitive skills have changed in time, and whether this can explain the concurrent increase in economic inequality.

Table 1: Item Commonalities

			BCS Rutter A Scale					MCS Strength and Difficulties				
Item	Factor	N. categ	Text		Cert. Applies	Somet. Applies	Applies	Text		Cert. True	Somet. True	True
1	EXT	3	Very restless. Often running about or jumping up and down. Hardly ever still	1	28.7	40.4		Restless, overactive, cannot stay still for long	2	13.7	25.7	
2	EXT	3	Is squirmy or fidgety	2	11.8	32.0		Constantly fidgeting or squirming	10	9.8	26.5	
3	EXT	3	Frequently fights other children + Bullies other children	4 19	4.9	33.9		Often fights with other children or bullies them	12	1.2	6.6	
4	EXT	3	Cannot settle to anything for more than a few moments	15	7.1	27.9		Easily distracted, concentration wanders	15	12.0	41.0	
5	EXT	2	Has temper tantrums	D			23.1	Often has temper tantrums or hot tempers	5			48.0
6	EXT	2	Not much liked by other children	5			5.9	(+) Generally liked by other children	14			12.0
7	EXT	2	Is often disobedient	14			69.4	(+) Generally obedient, usually does what adults request	7			44.3
8	INT	3	Often worried, worries about many things	6	5.7	30.3		Many worries, often seems worried	8	1.9	12.0	
9	INT	3	Tends to be fearful or afraid of new things or new situations	16	6.8	29.6		Nervous or clingy in new situations, easily loses confidence	16	10.2	36.2	
10	INT	3	Tends to do things on his/her own – rather solitary	7	9.1	36.4		Rather solitary, tends to play alone	6	5.1	24.3	
11	INT	3	Often appears miserable, unhappy, tearful or distressed	9	2.6	20.3		Often unhappy, down-hearted or tearful	13	1.3	7.5	
12	INT	2	Complains of headaches + Complains of stomach-ache or has vomited	A B			14.0	Often complains of head- aches, stomach-ache or sickness	3			18.4

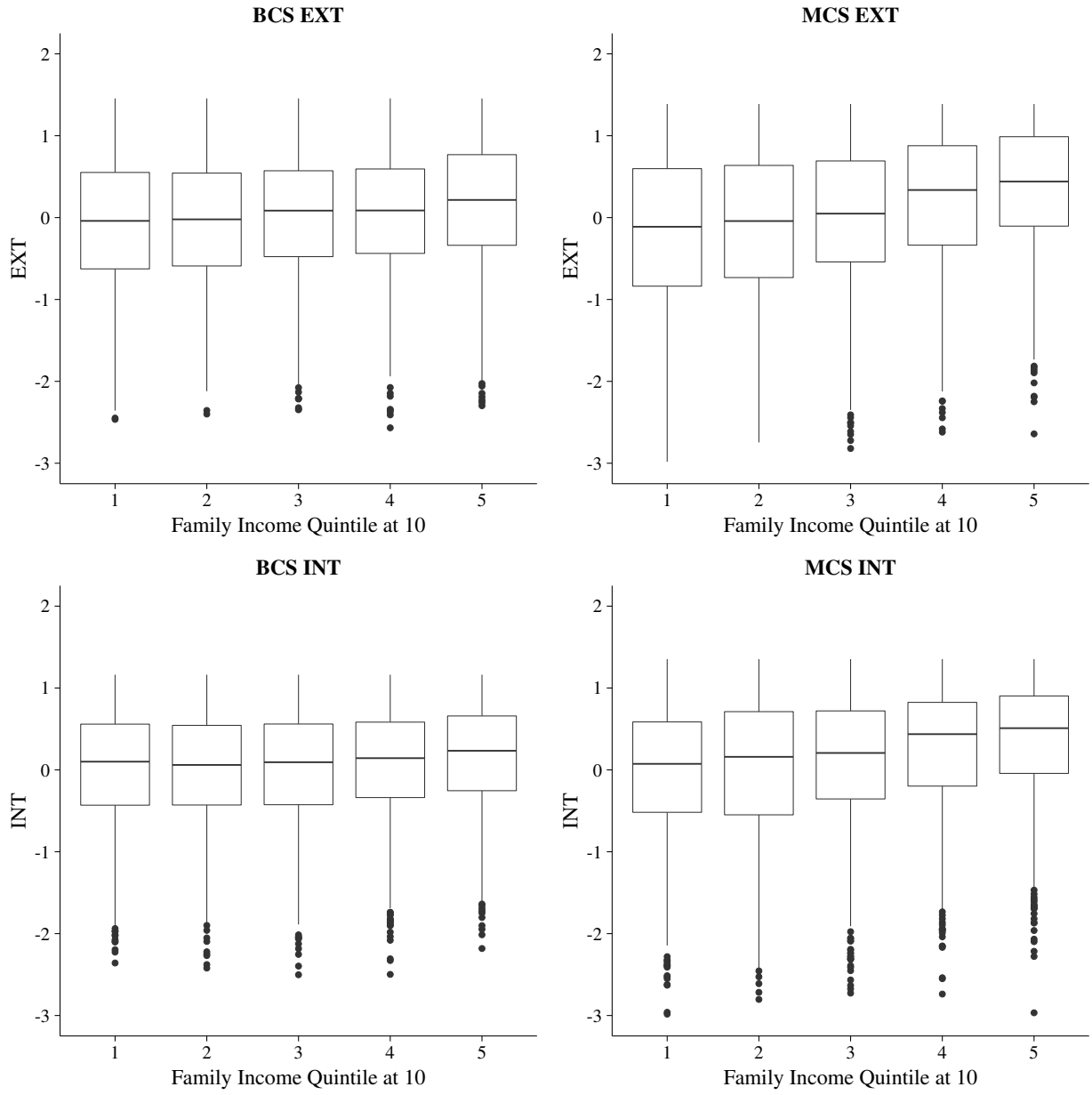


Figure 1: Distribution of scores across income quintiles

Notes: The figure is a boxplot of the distribution of the scored factors – externalising problems (*EXT*) and internalising problems (*INT*) – against the quintiles of the weekly family income distribution.

1 Bibliography

Blanden, J., P. Gregg, and L. Macmillan (2007, mar). Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education. *The Economic Journal 117*(519), C43–C60.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry and Allied Disciplines 36*(1), 113–149.

Carneiro, P., C. Crawford, and A. Goodman (2007). The impact of early cognitive and non-cognitive skills on later outcomes. *IFS working paper*, 1–39.

Cunha, F. and J. J. Heckman (2007). The Technology of Skill Formation. *AEA Papers and Proceedings 97*(2).

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, may). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Dickey, W. C. and S. J. Blumberg (2004, sep). Revisiting the Factor Structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry 43*(9), 1159–1167.

Goodman, A., D. L. Lamping, and G. B. Ploubidis (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the strengths and difficulties questionnaire (SDQ): Data from british parents, teachers and children. *Journal of Abnormal Child Psychology 38*(8), 1179–1191.

Goodman, R. (1994). A Modified Version of the Rutter Parent Questionnaire Including Extra Items on Children's Strengths: A Research Note. *Journal of Child Psychology and Psychiatry 35*(8), 1483–1494.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry 38*(5), 581–6.

Goodman, R., A. C. Iervolino, S. Collishaw, A. Pickles, and B. Maughan (2007). Seemingly minor changes to a questionnaire can make a big difference to mean scores: A cautionary tale. *Social Psychiatry and Psychiatric Epidemiology 42*(4), 322–327.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464.

Heckman, J. J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, jul). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). Fostering and Measuring Skills. (110).

Millsap, R. E. and J.-Y. Tein (2004, jul). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research 39*(3), 479–515.

Plewis, I. (2007). Millennium Cohort Study: Technical Report on Sampling. Technical Report July.

Rutter, M., J. Tizard, and K. Whitmore (1970). *Education, health, and behaviour*. London, UK: Prentice Hall Press.

Segal, C. (2008). Classroom Behavior. *The Journal of Human Resources 43*(4), 783–814.

Stone, L. L., R. Otten, R. C. Engels, A. A. Vermulst, and J. M. Janssens (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-Year-olds: A review. *Clinical Child and Family Psychology Review 13*(3), 254–274.

Wu, H. and R. Estabrook (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika 81*(4), 1014–1045.