EDUCATION & SKILLS

Beyond self-reports: International evidence on parent and teacher participation in assessing students' social and emotional skills

A multi-stage analysis of multi-informant data from SSES 2019 (Round 1)

Diego Maury Romero, (former) OECD SSES Team

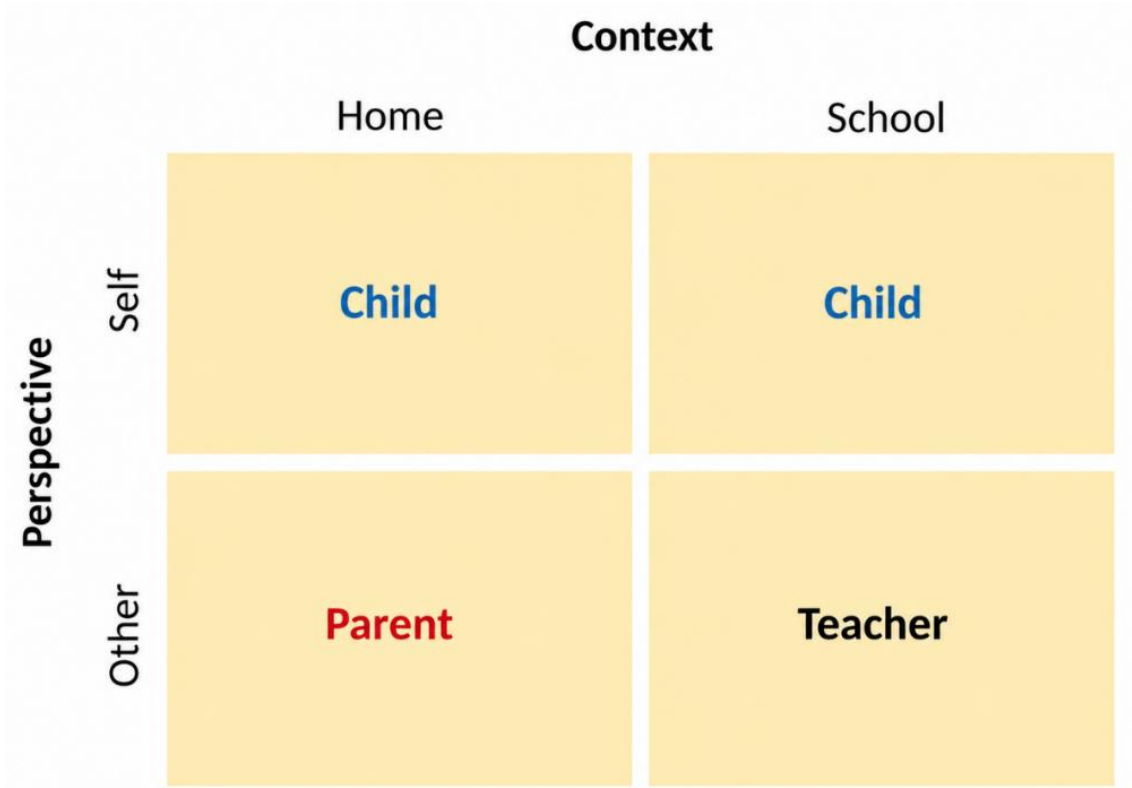
SSES Seminar
02-Jul-2026

OECD
BETTER POLICIES FOR BETTER LIVES

Background

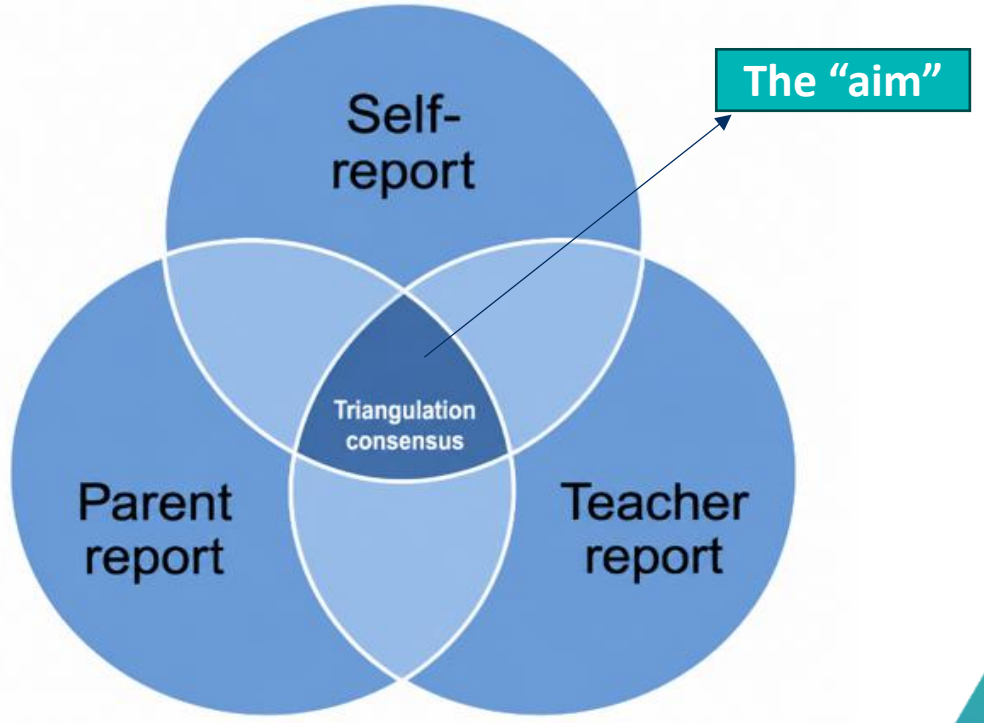
Triangulation in youth and education studies

Informant selection > number of informants



Note. Figure adapted from Kraemer et al. (2003)

Multi-informant triangulation



Restricted Use - À usage restreint

Background

Triangulation in SSES 2019

Assessment framework

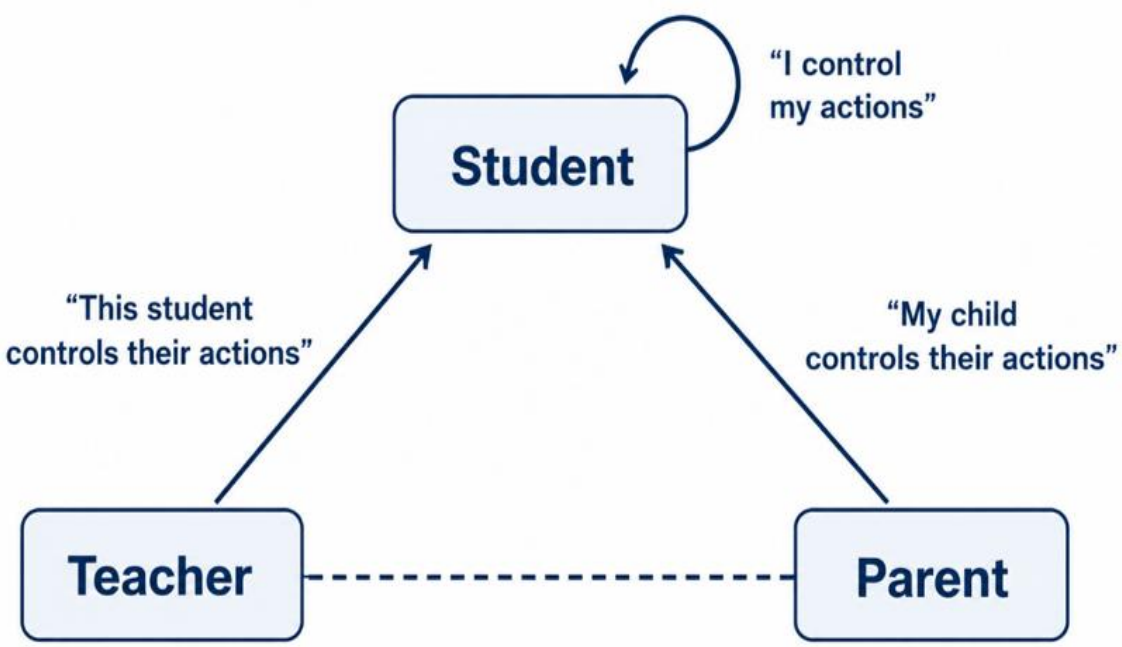
SSES 2019 maps **Big Five domains** onto education-facing domains:

- Open-mindedness,
- Task performance,
- Emotional regulation,
- Collaboration,
- Engaging with others (*Social engagement*).

Each domain comprises **three skills**.

Each construct is measured by **three items**, developed for students and adapted to teacher and parent applications.

Multi-informant design in SSES



Note: The informants report on the same Likert-type scale with five response choices ranging from "Strongly disagree" to "Strongly agree".

Restricted Use - À usage restreint

The Present Study

A systematic multi-stage analytical approach to triangulation of multi-informant reports

Research Questions

- **RQ1.** Do SSES student, parent and teacher reports support a valid multi-informant MTMM structure for selected social and emotional skills?

  Construct validity

- **RQ2.** How is variance in the selected SSES skill measures distributed across student-reference, informant-specific and residual components?

  Convergence analysis

- **RQ3.** How do parent and teacher reports contribute to triangulated skill assessment beyond student self-reports?

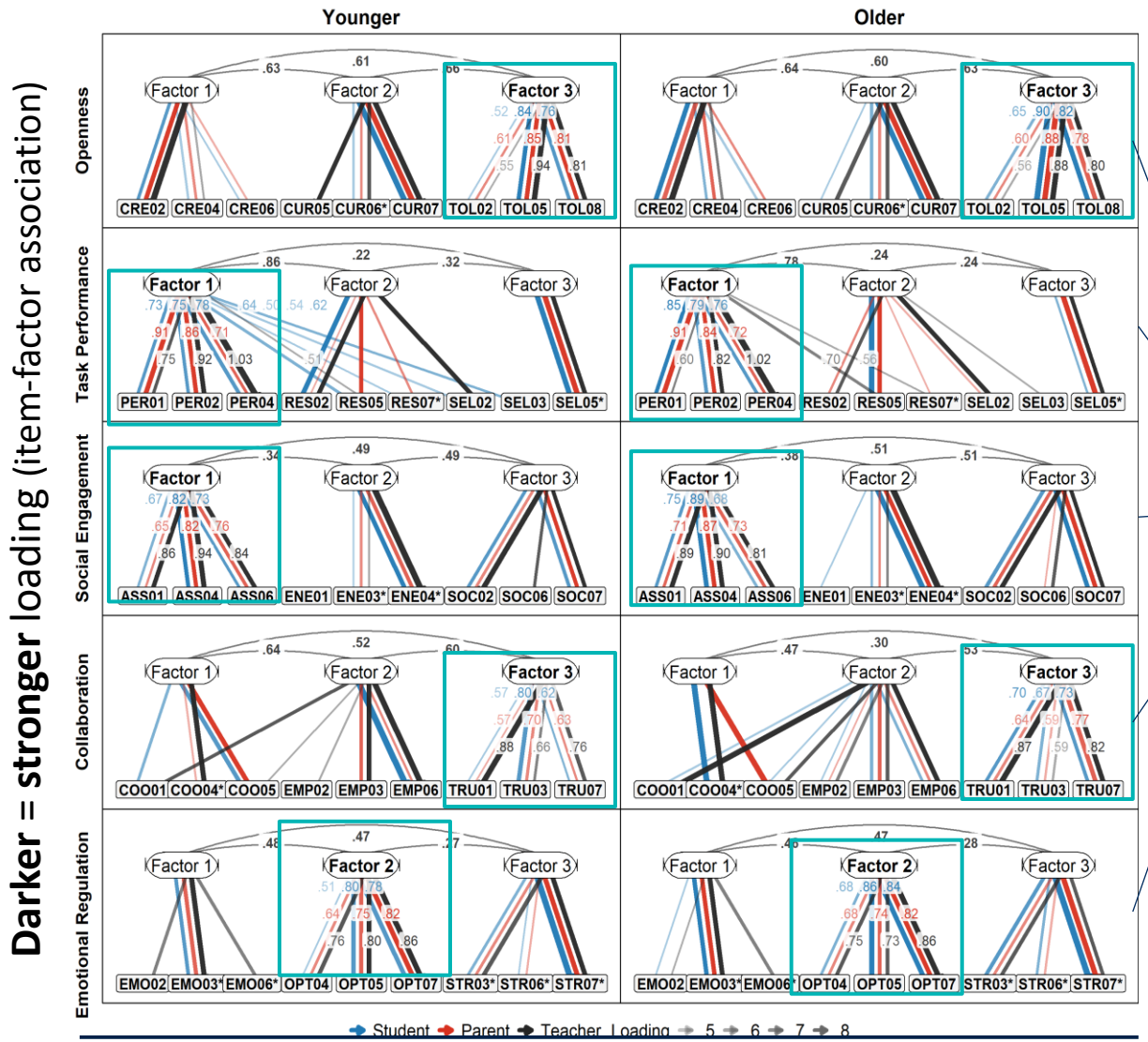
  Consensus analysis

Analytical stages

Stage	Aim	Method	Main output
1	Comparison of *competing* hypotheses *Convergent* validity **+** *Discriminant* validity **=** Construct validity	**CFA-MTMM** ***Trait*** = skill ***Method*** = inform	ΔCFI
2	Comparison of parent and teacher reports with student self-reports Reference-method sensitivity testing	**Multiple-indicator CT-C(M-1) model**	Variance shares (%)
3	Comparison of precision and information contributions	**Bayesian rater-facet GRM** ***Rater*** = informant	Precision weights

Restricted Use - À usage restreint

What skills share a robust measurement structure?



One skill survived in each domain.

The IRT-EFA screening identifies a set of **5** skills with sufficiently high (> .5) item-factor loading across *all three* informants:

- ***Tolerance*** (TOL) from Openness.
- ***Persistence*** (PER) from Task Performance.
- ***Assertiveness*** (ASS) from Social Engagement.
- ***Trust*** (TRU) from Collaboration.
- ***Optimism*** (OPT) from Emotional Regulation.

Each informant-skill combination has **high empirical reliability**, *rxx* ≥ 0.7.

Restricted Use - À usage restreint

Multi-informant reports provide coherent skill information

The evaluated multi-informant reports support triangulated and distinct skill measurements.

CFA-MTMM Model Comparisons

CT-CM = traits and methods correlate *freely* with each other. This is the benchmark model used for comparison (H_1).

Gains in fit ($\Delta CFI \geq .01$) = **support** for H_1.

Cohort	Test (H0)	Model compared	$\Delta\chi^2$	Δdf	ΔCFI	$\Delta RMSEA$	Reading
Younger (CFI = .95)	Convergent validity	NT-CM	20779.33	55	0.122	0.046	**Supported**
	Discriminant validity	PCT-CM	8209.11	10	0.072	0.032	**Supported**
	Method correlation	CT-UM	104.59	3	-0.003	-0.002	Not supported
Older (CFI = .94)	Convergent validity	NT-CM	27271.82	55	0.166	0.051	**Supported**
	Discriminant validity	PCT-CM	8857.28	10	0.112	0.040	**Supported**
	Method correlation	CT-UM	73.23	3	-0.007	-0.004	Not supported

NT = No traits (only methods). **PCT =** Perfectly-correlated (equal) traits.
UM = Un-correlated (independent) methods.

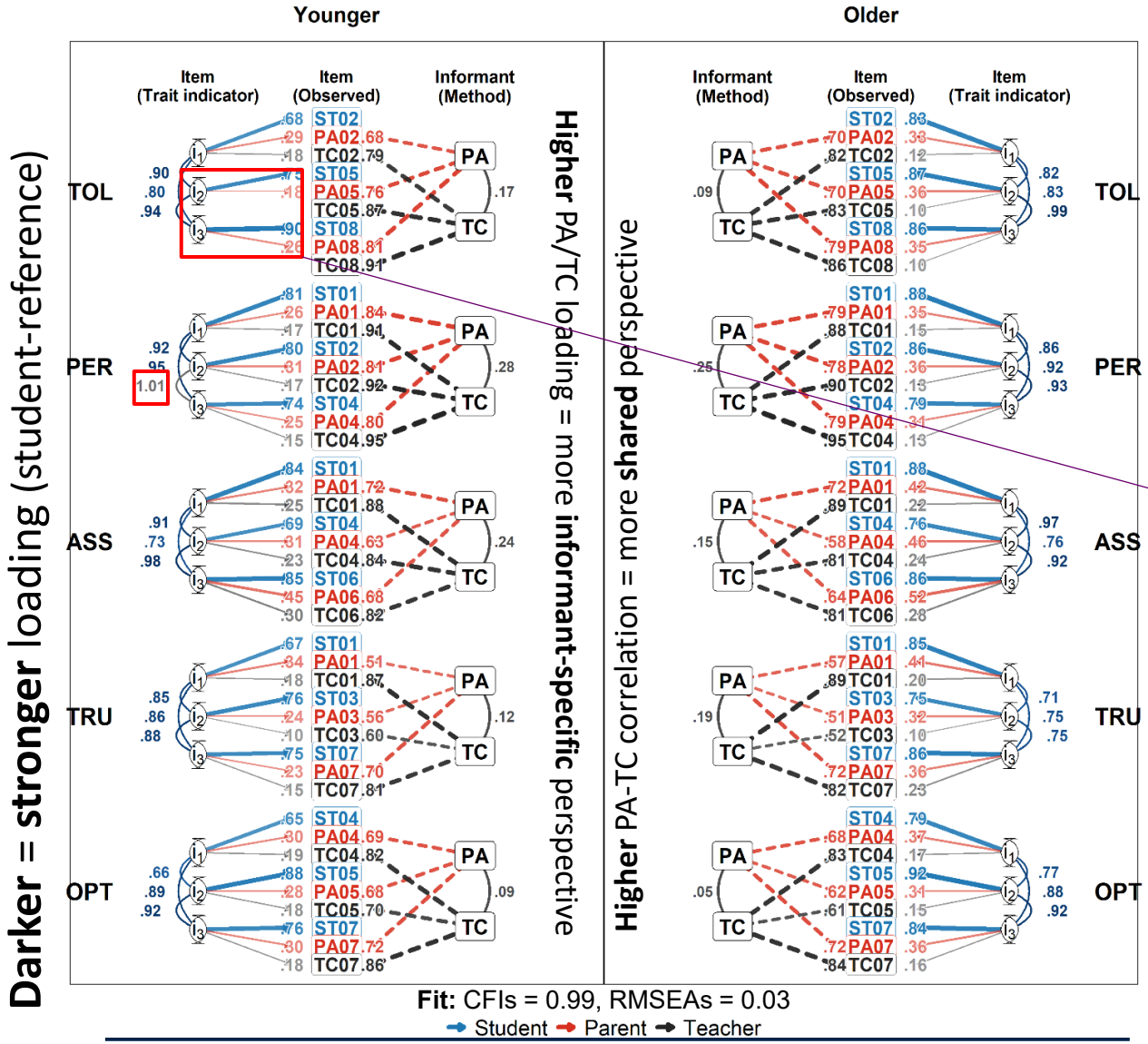
Key Findings

- Convergent validity is **supported** in both cohorts.
- Discriminant validity is **supported** in both cohorts.
- Accounting for *latent* associations between informants does **not** help improve multi-informant model fit.

The retained measures can therefore be analysed as a coherent, independent multi-informant measurement system.

Restricted Use - À usage restreint

Adult reports *moderately* converge with student self-reports



Student-reference convergence is limited

Adult reports share *some* information with student self-reports, but:

- Parent and teacher answers have **weak-to-moderate** links to student answers.
- Convergence is generally higher for the Older cohort than the Younger cohort.
- Tolerance (TOL) shows unstable (non-significant) item comparability for Younger students.
- Teacher (TC) reports' specificity is very high (>.8), higher than parents' in every skill.

Parent reports have more in common with student self-reports than teacher reports. They may also provide a more reliable baseline for CT-CM(M-1) models of younger-student skills.

(A parent-reference model showed no non-significant loadings)

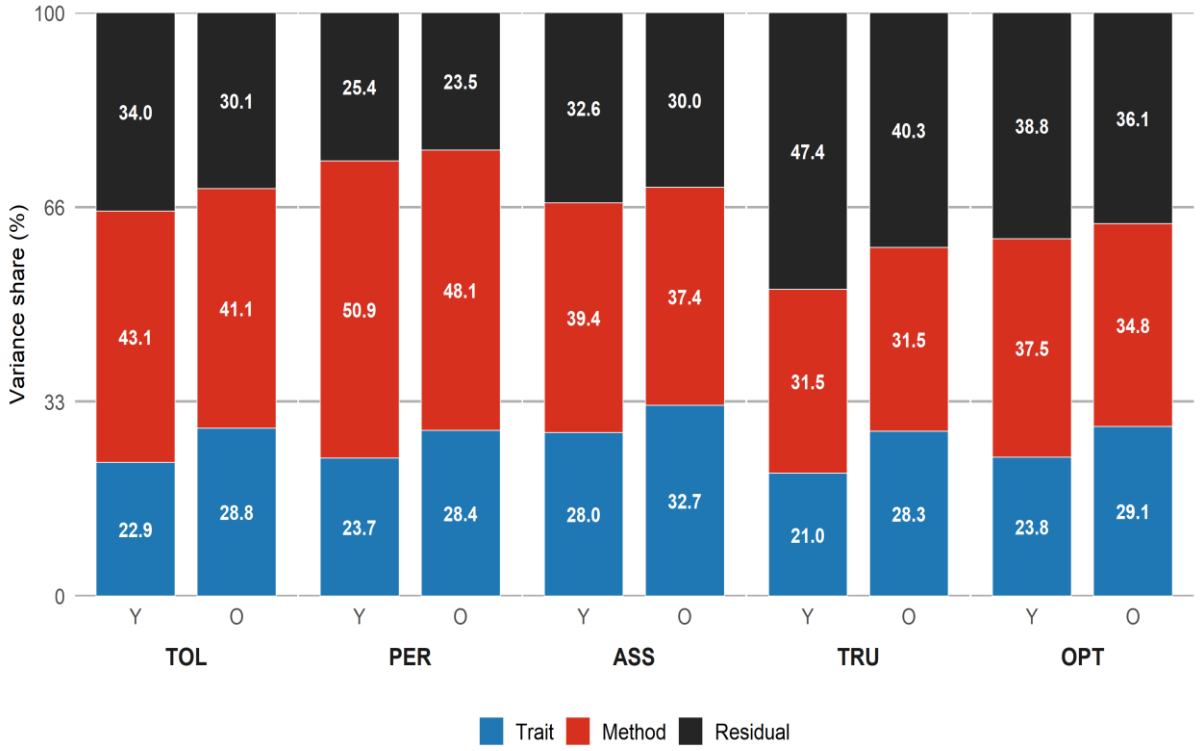
Restricted Use - À usage restreint

Triangulation capacity is often limited but varies with age and measure

Triangulation works best when measuring students' social and performance-related skills.

Response variance breakdown

Less than a third can be attributed to a **shared skill** measure; most is between-informant differences.



Only **three** skills show acceptable reliability

- **Tolerance (TOL)** shows *acceptable* reliability (66 pp [Y] - 70 pp [O]).
- **Perseverance (PER)** is **the most reliable** (75 - 76) but mainly due to informant specificity.
- **Assertiveness (ASS)** is as reliable as TOL (0.67 - 0.70) but with higher **convergence**, 0.36 [Y] to 0.47 [O].

Trust (TRU) and Optimism (OPT) are therefore dropped from subsequent analyses.

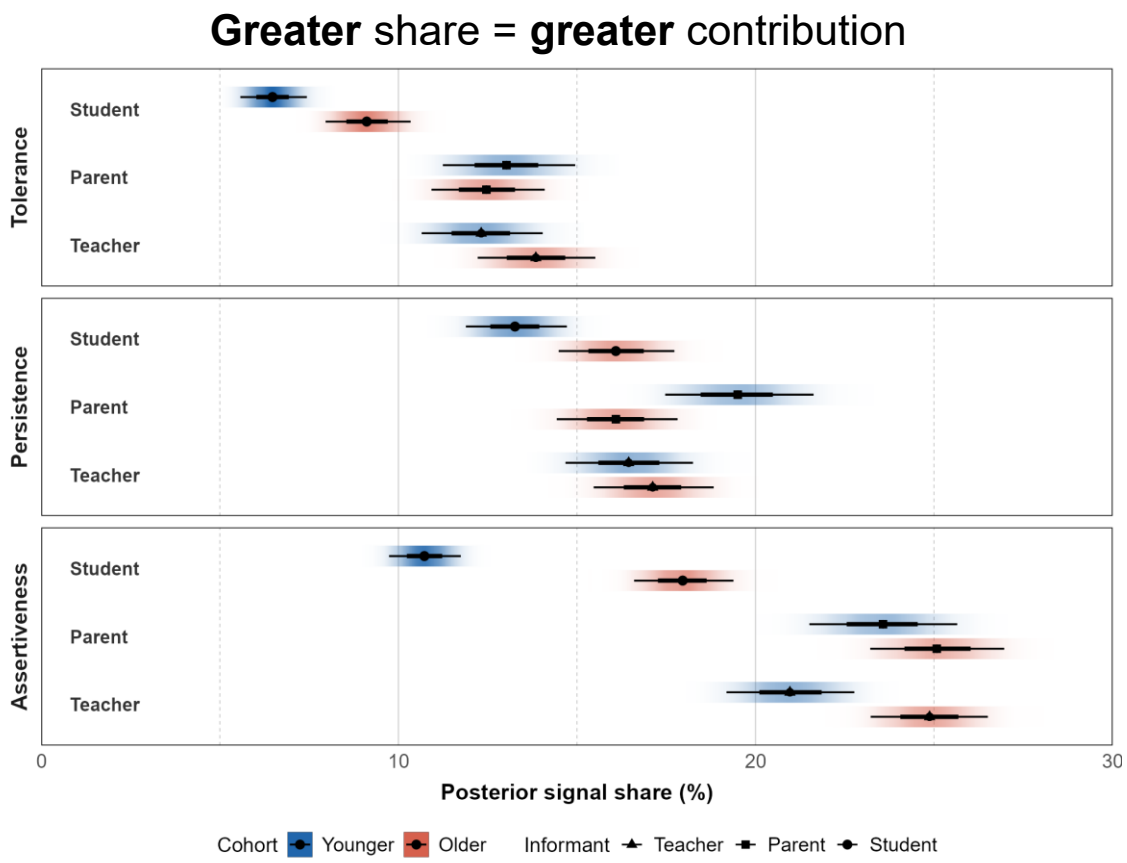
Older-student (O) reports *consistently* provide more comparable estimates.

Restricted Use - À usage restreint

Multilevel GRMs provide a robust alternative to CFA/SEM frameworks

A Multilevel Graded Response Model (GRM) provides comparable yet richer results than CT-C(M-1) models

Contributions to cross-informant concensus



Shared-measurement is strongest for Assertiveness

The skill-signal overlap is strongest for Assertiveness and the weakest for Tolerance.

- Tolerance shows the weakest triangulation capacity. **Students contribute little**, while adults have comparable, higher shares.
- Assertiveness has the clearest shared signal. **Younger students contribute much less** than their parents and their teachers.
- Persistence has the most comparable shares. **Parents of younger students contribute remarkably more** than other informants.

Fit summaries = R-hats < 1.01; bulk and tail ESS > 300

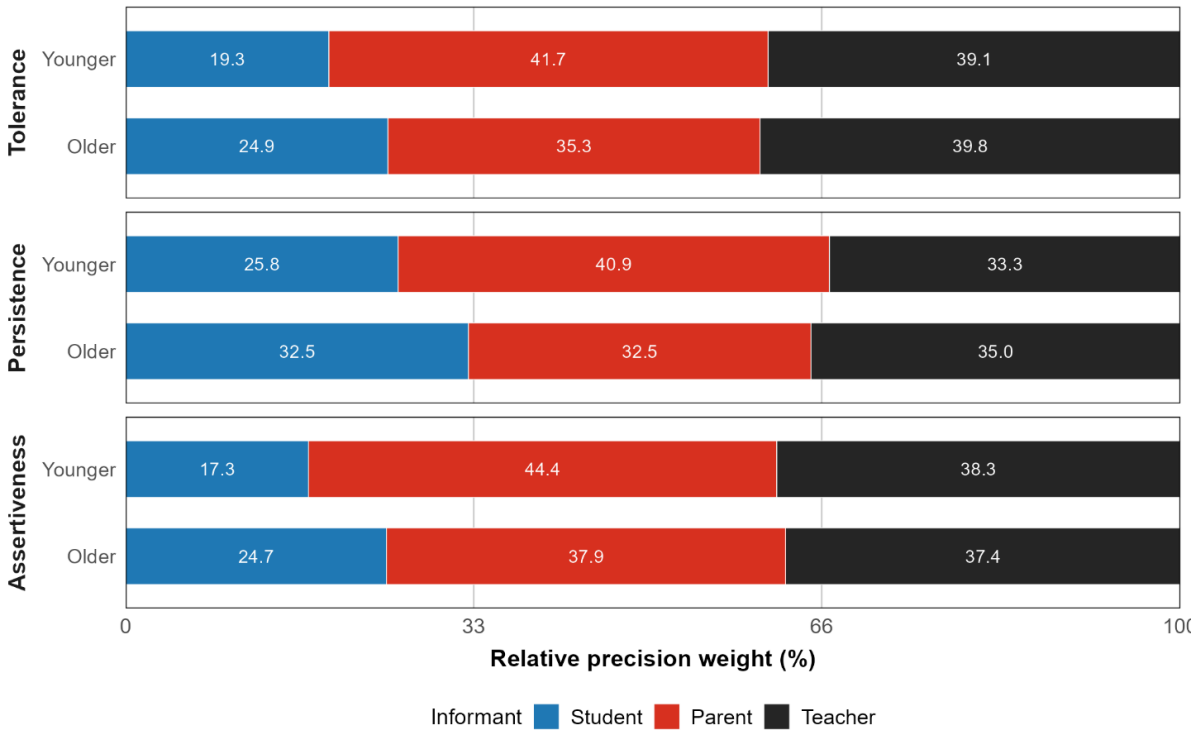
Restricted Use - À usage restreint

Triangulation appears most valuable for younger-student assessments

15-year-olds provide, on average, more precise information than 10-year-olds

Relative contributions to cross-informant concensus

Higher weight = **more** precise information



Precision differs by age and skill measure

- Student self-reports tend to be noisier than that adult reports, contributing **less than one-third** of triangulation precision.
- **Parents contribute especially strongly** to multi-informant consensus for Younger student assessments.
- **Teachers provide** consistent amounts of precision—regardless of skill—although slightly more than parents for Tolerance and Persistence among older students.

Adult reports could help compensate for weak student self-report measures.

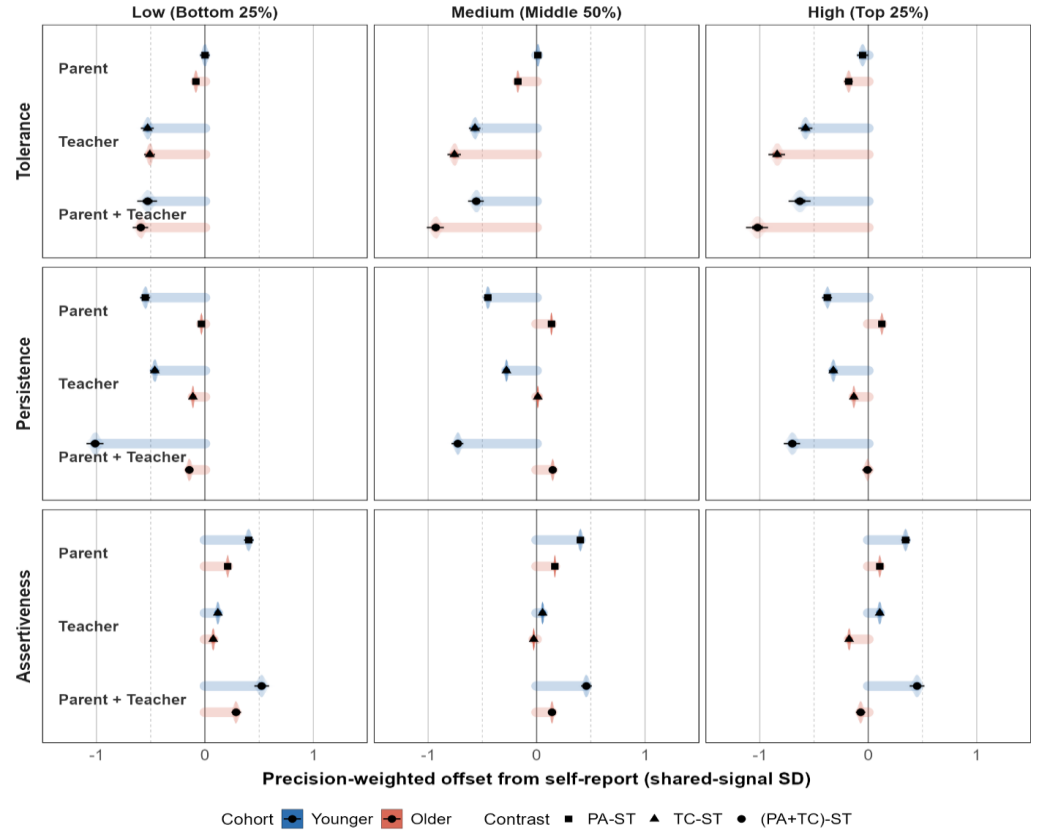
Restricted Use - À usage restreint

Adult reports also contribute qualitatively distinct information

Triangulation could help offset ceiling effects and add meaningful perspectives.

Direction contributions of incorporating adult reports

Upward pull = **more positive**. **Downward = more**



The offset effect varies by skill and, sometimes, skill level

- Tolerance shifts skill estimates **down** from the student baseline, especially relative to teachers and for older students.
- Perseverance student estimates are also pulled down for younger students, most notable towards the lower end.
- Assertiveness shows an **upward** adult-informant pull, especially for younger student assessments.

Triangulation might help adjust extremely positive or negative skill representations.

Restricted Use - À usage restreint

Conclusions

Multi-informant skill measurement is empirically achievable, but its value varies across measures, age cohorts, and informants.

RQ1–Construct Validity

SSES constructs *can* support a valid multi-informant measurement structure across student, parent, and teacher reports.

RQ2–Measure Convergence

Convergence with student self-reports is higher for 15-year-olds than for 10-year-olds and with parents than with teachers.

RQ3–Consensus Breakdown

Adult reports increase measurement precision and, in some cases, lead to systematically different estimates compared to students.

Future research

Further research is needed to understand the determinants of convergence and the *incremental* validity of these triangulations.

Restricted Use - À usage restreint


EDUCATION
& SKILLS
Questions? Thoughts?


EDUCATION
& SKILLS
Thank you!