2.5 Modelling equates

Iris Eekhout[1]
Stef van Buuren[1,2]
[1]Netherlands Organisation for Applied Scientific Research TNO, Leiden, 2316 ZL, The Netherlands
[2]University of Utrecht, Utrecht, 3584 CH, The Netherlands

This section deals with the nitty-gritty of the modelling strategy used for the GCDG data introduced in Section 2.2. This section

- provides a high-level description of the GCDG data (2.5.1)
- discusses various modelling strategies (2.5.2)
- shows the impact of equate groups on the model in extreme cases (2.5.3)
- demonstrates visualization of age profiles to select promising equate groups (2.5.4)
- introduces a helpful visualization of the quality of the equate group (2.5.5)
- highlights infit and outfit for removing misfitting milestones (2.5.6)
- discusses instrument fit and equate group editing (2.5.7)
- introduces a grading system for equate groups (2.5.8)
- provides pointers to the final model (2.5.9)

2.5.1 GCDG DATA: DESIGN AND DESCRIPTION

2.5.1.1 Data combination

Section 2.2.1 provides an overview of the data collected by Global Child Development Group. The group collected item level measurements obtained on 12 instruments for measuring child development across 16 cohorts.

We coded every item as 0 (FAIL), 1 (PASS) or missing. For some instrument we did some additional recoding to restrict to these two response categories. The Battelle Developmental Inventory scores items as 0 (FAIL), 1, or 2, depending on the level of skill demonstrated or time taken to complete the task. We joined categories 1 and 2 for these items. The ASQ items were originally scored as 0 (not yet), 5 (sometimes) and 10 (succeeds). We recoded both 5 and 10 to 1.

We concatenated the datasets from the GCDG cohorts cohort. The resulting data matrix has 71403 rows (child-visit combinations) and 1572 columns (items) collected from 36345 unique children. We removed 233 items that had fewer than 10 observations in a category. The remaining 1339 items were candidates for analysis. The total number of observed scores was equal to about

DOI: 10.1201/9781003216315-16

2.8 million pass/fail responses. While this is a large number of measurements, about 97 percent of the entries in the matrix are missing.

2.5.1.2 Equate group formation

A group of 13 subject-matter experts from the Global Child Development Group cross-walked the available instruments for similar milestones. This group

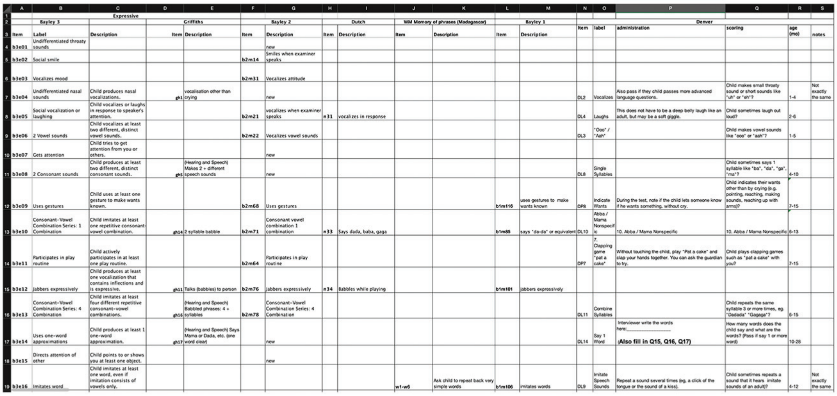
- developed an item coding schema;
- matched similarly appearing items stemming from different instruments;
- formed an opinion about the quality of each match;
- noted peculiarities of the matches;
- reported the results as a series of detailed Excel spreadsheets.

The group evaluated around 1500 milestones. After several days, this highly-skilled, intensive labour resulted in a series of spreadsheets. Figure 2.5.1 shows an example. These sheets formed the basis of an initial list of 184 equate groups, each consisting of at least two items.

2.5.2 MODELLING STRATEGIES

The analytic challenge is twofold:

- to find a subset of items that form a scale;
- to find a subset of equate groups with items similar enough to bridge instruments.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		Expressive																	
2		Bayley 3			Griffiths		Bayley 2		Dutch		WM Memory of phrases (Madagascar)		Bayley 1			Denver			
3	Item	Label	Description	Item	Description	Item	Description	Item	Description	Item	Description	Item	Description	Item	label	administration	scoring	age (mo)	notes
4	b3e01	Undifferentiated throaty sounds					new												
5	b3e02	Social smile				b2m14	Smiles when examiner speaks												
6	b3e03	Vocalizes mood				b2m31	Vocalizes attitude												
7	b3e04	Undifferentiated nasal sounds	Child produces nasal vocalizations.	gh1	vocalisation other than crying		new							DL2	Vocalizes	Also pass if they child passes more advanced language questions.	Child makes small throaty sound or short sounds like "uh" or "eh"?	1-4	Not exactly the same
8	b3e05	Social vocalization or laughing	Child vocalizes or laughs in response to speaker's attention.			b2m21	vocalizes when examiner speaks	n31	vocalizes in response					DL4	Laughs	This does not have to be a deep belly laugh like an adult, but may be a soft giggle.	Child sometimes laugh out loud?	2-6	
9	b3e06	2 Vowel sounds	Child vocalizes at least two different, distinct vowel sounds.			b2m22	Vocalizes vowel sounds							DL3	"Ooo" / "Aah"		Child makes vowel sounds like "ooo" or "aah"?	1-5	
10	b3e07	Gets attention	Child tries to get attention from you or others.				new												
11	b3e08	2 Consonant sounds	Child produces at least two different, distinct consonant sounds.	gh5	(Hearing and Speech) Makes 2 + different speech sounds		new							DL8	Single Syllables		Child sometimes says 1 syllable like "ba", "da", "ga", "ma"?	4-10	
12	b3e09	Uses gestures	Child uses at least one gesture to make wants known.			b2m68	Uses gestures					b1m116	uses gestures to make wants known	DP8	Indicate Wants	During the test, note if the child lets someone know if he wants something, without cry.	Child indicates their wants other than by crying (e.g. pointing, reaching, making sounds, reaching up with arms)?	7-15	
13	b3e10	Consonant-Vowel Combination Series: 1 Combination	Child imitates at least one repetitive consonant-vowel combination.	gh14	2 syllable babble	b2m71	Consonant vowel combination 1 combination	n33	Says dada, baba, gaga			b1m85	says "da-da" or equivalent	DL10	Abba / Mama Nonspecific	10. Abba / Mama Nonspecific	10. Abba / Mama Nonspecific	6-13	
14	b3e11	Participates in play routine	Child actively participates in at least one play routine.			b2m64	Participates in play routine							DP7	7. Clapping game "pat a cake"	Without touching the child, play "Pat a cake" and clap your hands together. You can ask the guardian to try.	Child plays clapping games such as "pat a cake" with you?	7-15	
15	b3e12	Jabbers expressively	Child produces at least one vocalization that contains inflections and is expressive.	gh11	Talks (babbles) to person	b2m76	Jabbers expressively	n34	Babbles while playing			b1m101	jabbers expressively						
16	b3e13	Consonant-Vowel Combination Series: 4 Combination	Child imitates at least four different repetitive consonant-vowel combinations.	gh16	(Hearing and Speech) Babbled phrases: 4 + syllables	b2m78	Consonant-Vowel Combination Series: 4 Combination							DL11	Combine Syllables		Child repeats the same syllable 3 or more times, eg. "Dadada" "Gagaga"?	6-15	
17	b3e14	Uses one-word approximations	Child produces at least 1 one-word approximation.	gh17	(Hearing and Speech) Says Mama or Dada, etc. (one word clear)		new							DL14	Say 1 Word	Interviewer write the words here:____________________ (Also fill in Q15, Q16, Q17)	How many words does the child say and what are the words? (Pass if say 1 or more word)	10-28	
18	b3e15	Directs attention of other	Child points to or shows you at least one object.				new												
19	b3e16	Imitates word	Child imitates at least one word, even if imitation consists of vowels only.							w1-w6	Ask child to repeat back very simple words	b1m106	imitates words	DL9	Imitate Speech Sounds	Repeat a sound several times (eg, a click of the tongue or the sound of a kiss).	Child sometimes repeats a sound that it hears imitate sounds of an adult)?	4-12	Not exactly the same

FIGURE 2.5.1 A snapshot of information generated by subject-matter experts.

Note that both subsets are related, i.e., changing one affects the other. Thus, we cannot first identify items and then equate groups, or first identify equate groups followed by the items. Rather we need to find the two subsets in an iterative fashion, primarily by hand. This section describes some of the modelling issues the analyst needs to confront.

In general, we look for a final model that

- preserves the items that best fit the Rasch model;
- uses active equate groups with items that behave the same across many cohorts and instruments;
- displays reasonable age-conditional distributions of the D-scores;
- has difficulty estimates that are similar to previous estimates.

The modelling strategy is a delicate balancing act to achieve all of the above objectives. Particular actions that we could take to improve a given model are:

- remove bad items;
- inactivate bad equate groups;
- break up bad equate groups;
- move items from one equate group to another;
- create new equate groups;
- remove entire instruments;
- remove persons;
- remove studies.

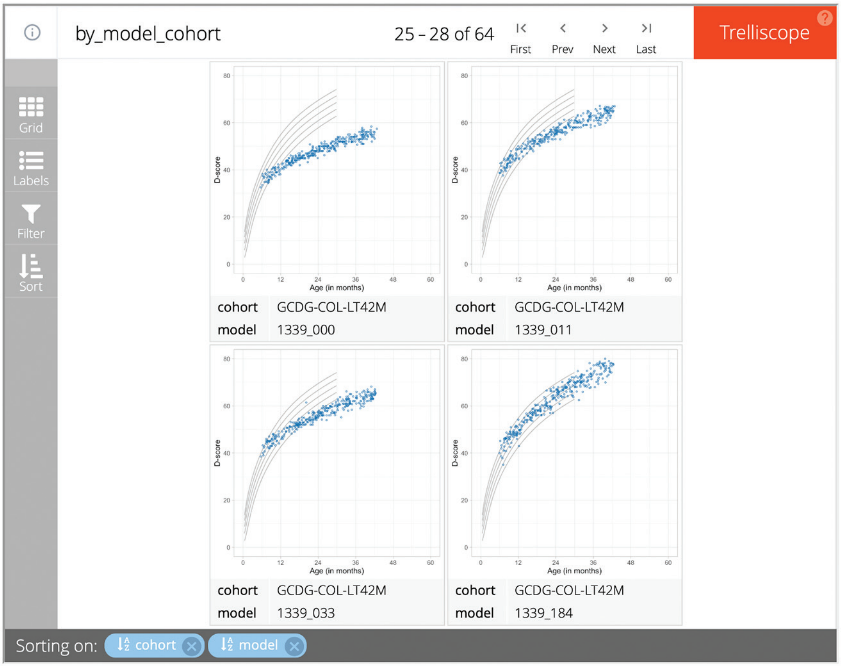
In order to steer our actions, we look at the following diagnostics (in order of importance):

- quality of equate groups (both visually and through infit);
- plausibility of the distribution of the D-score by age per study;
- correspondence of difficulty estimates from published (single study) Dutch data and the new model;
- infit of the items remaining in the model.

Various routes are possible and may result in different final models. The strategy adopted here is to thicken active equate groups by covering as many studies as possible, in the hope of minimizing the number of active equates needed.

2.5.3 IMPACT OF NUMBER OF ACTIVE EQUATE GROUPS

Figure 2.5.2 is a display of the D-score by age for the `GCDG-COL-LT42M` cohort under four models. D-score by age visualizations for all cohort are can be found via this link. As a rough reference to compare, the grey curves in the back represent the Dutch model as calculated from the SMOCC study. In order



cohort GCDG-COL-LT42M
model 1339_000

cohort GCDG-COL-LT42M
model 1339_011

cohort GCDG-COL-LT42M
model 1339_033

cohort GCDG-COL-LT42M
model 1339_184

FIGURE 2.5.2 D-score by age of four models with all 1339 items using 0, 11, 33 and 184 active equate groups.

The number of equate groups has a substantial effect on the D-score distribution (https://d-score.org/dbook-apps/models1339/, in the online version you can use the arrows to see other cohorts).

to speed up the calculations, the figure shows a random subsample of 25% of all points. Manipulate the plot controls to switch cohorts.

All models contain 1339 items, but differ in the number of active equate groups. The most salient features per model are:

- `1339_`: No equate groups, so different instruments in different cohorts are fitted independently;
- `1339_11`: Connects all cohorts through one or more equated items using 11 equate groups in total;
- `1339_33`: There are 33 equate groups that bridge cohort and instruments;
- `1339_184`: Maximally connects instruments and cohort by all equate groups.

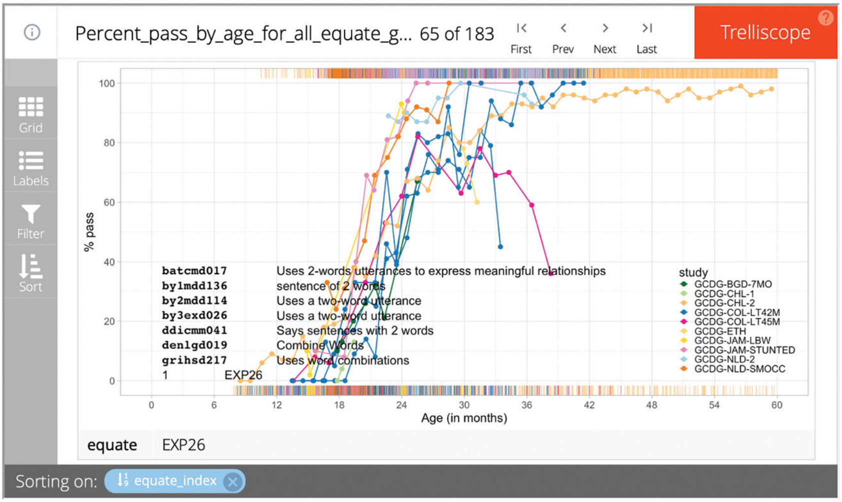
Comparison of the D-score distribution by age across these models yields various insights:

- The location of cohorts on the vertical scale depends on the number of active equate groups. For example, for Madagascar (MDG) the points are located around 52 when no equate groups are activated, whereas if all are activated it is about 68.
- The age trend depends on the number of active equate groups. For example, for Colombia (COL) or Ethiopia (ETH), the model without equate groups has a shallow age trend, whereas it is steep for the `1339_184` model.
- The vertical spread depends on the number of equate groups. For example, the spread in the Chile-2 (CHL-2) cohort substantially increases with the number of active equates.
- Model `1339_0` for the Dutch NLD-SMOCC cohort is equivalent to the model fitted to the SMOCC study alone. Introducing equate groups compresses the range of scores, especially at the higher end.

We have now seen that the number of active equate groups has a large effect on the model. The next sections look into the equate groups in more detail.

2.5.4 AGE PROFILES OF SIMILAR MILESTONES

Figure 2.5.3 displays the percentage of children that pass milestones at various ages for equate group EXP 26. Subject matter experts clustered similar items stemming from different instruments into equate groups. There are 184 equate groups that contain two or more milestones; the percentage pass by age for the items in these equate groups are shown here.



batcmd017	Uses 2-words utterances to express meaningful relationships
by1mdd136	sentence of 2 words
by2mdd114	Uses a two-word utterance
by3exd026	Uses a two-word utterance
ddicmm041	Says sentences with 2 words
denlgd019	Combine Words
grihsd217	Uses word combinations
1	EXP26

FIGURE 2.5.3 Percentage of children that pass similar milestones at a given age (https://d-score.org/dbook-apps/p-a-equate-1339/).

Most age profiles show a rising pattern, as expected, though some (e.g. FM17 or EXP11) have one item showing a negative relation with age. Equate EXP26 combines two-word sentences items from seven instruments into one plot. The item difficulties expressed as age-equivalents (cf. Section 1.3.1.2, Chapter I (van Buuren & Eekhout, 2021)) for these cohorts vary between 20–25 months. By comparison, equate group EXP18 (says two words) shows more heterogeneity across cohorts, and is therefore, less likely to be useful for equating. Equate group FM31 (stack two blocks) is another example of a promising example. By comparison, FM38 (stack 68 blocks) shows additional heterogeneity. As a last example, consider GM42 (walks alone), which has a similar age profile across cohorts, whereas GM44 (throws ball) or GM49 (walk down stairs) are more heterogeneous.

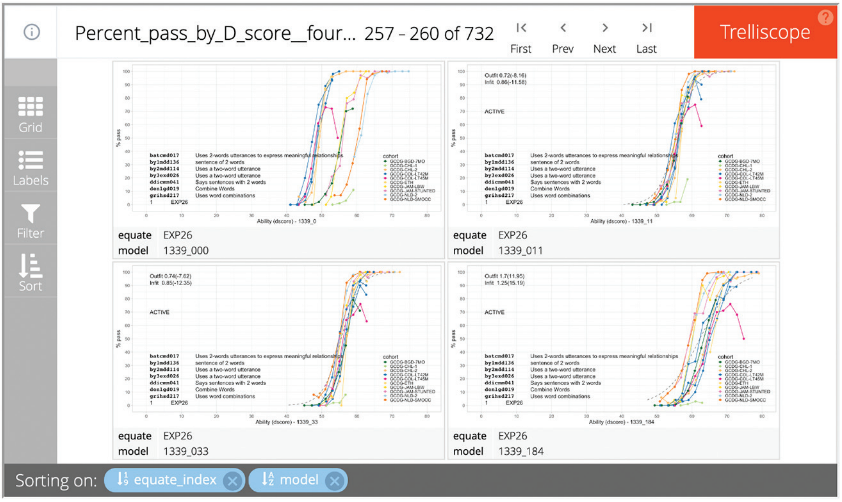
We could follow different strategies in selecting which equate groups to activate. One strategy would be to include as many equate groups as possible (e.g. all 184 equates) so as to build as many bridges as possible between different instruments. A more selective strategy would be to activate a subset of promising equates and leave others inactive. The following section compares four different approaches.

2.5.5 QUALITY OF EQUATE GROUPS

This visualization shows how the passing percentage depends on the child's D-score as calculated under four models. All models include the same 1339 milestones, but differ in the number of active equates. The grey curve corresponds to the estimate made under the assumption that milestones are equally difficult. Good milestones for bridging instruments will have a tight bundle of curves. For example, as shown in Figure 2.5.4, equate EXP26 has tight bundles especially in models 1339_11 and 1339_33. By comparison, the curves of the two extreme models vary considerably: the model without any bridges (1339_) or the model with all bridges (1339_184) are thus less than ideal. The shallow grey curve of model 1339_184 indicates a poorer overall fit.

Outfit and infit statistics measure the residual deviation of the items to the grey curve. High values (e.g. above 1.4) are undesirable and indicate lack of fit to the model. For example, the fit statistics for EXP26 in model 1339_184 (1.70 and 1.25) indicate a mediocre fit, whereas EXP26 in models 1339_33 and 1339_11 fits well. Sometimes the individual item curves are steeper than the grey curve. This indicates that these milestones are more discriminative than the combined item. Model 1339_ lacks a grey curve and has no fit statistics for equate groups, because in that model, the combined item is not activated.

The probability curves provide a quick visual method for spotting promising and problematic equate groups. Examples of promising equate groups include COG36, FM31, GM26 and GM42. A little more weak are FM26 (has more variability), FM52 (looks promising, but has a problem with the item grigcd42 from the GCDG_JAM_STUNTED cohort), and GM35 (does not align cohort


Percent_pass_by_D_score__four... 257 – 260 of 732
First
Prev
Next
Last
Trelliscope
Grid
Labels
Filter
Sort
% pass
Outfit 0.72(-8.16)
Infit 0.86(-11.58)
Outfit 0.74(-7.62)
Infit 0.85(-12.35)
Outfit 1.7(11.95)
Infit 1.25(15.19)
ACTIVE
batcmd017 Uses 2-words utterances to express meaningful relationships
by1mdd136 sentence of 2 words
by2mdd114 Uses a two-word utterance
by3exd026 Uses a two-word utterance
ddicmm041 Says sentences with 2 words
denlgd019 Combine Words
grihsd217 Uses word combinations
1 EXP26
cohort
GCDG-BGD-7MO
GCDG-CHL-1
GCDG-CHL-2
GCDG-COL-LT42M
GCDG-COL-LT45M
GCDG-ETH
GCDG-JAM-LBW
GCDG-JAM-STUNTED
GCDG-NLD-2
GCDG-NLD-SMOCC
Ability (dscore) - 1339_0
Ability (dscore) - 1339_11
Ability (dscore) - 1339_33
Ability (dscore) - 1339_184
Sorting on: equate_index model

equate EXP26
model 1339_000

equate EXP26
model 1339_011

equate EXP26
model 1339_033

equate EXP26
model 1339_184

FIGURE 2.5.4 Percentage of children that pass similar milestones given their D-score as calculated under four models (1339 items, and 0, 11, 33 and 184 equate groups, respectively (https://d-score.org/dbook-apps/p-d-equate-1339/).

GCDG-ZAF). In such cases, one may wish to move an item out of an equate group, combine equate groups, or inactivate troublesome links.

Until now we only looked at models that include all 1339 items. In practice, we may improve upon the model by selecting the subset of milestones that fit the Rasch model. The next section looks in this modelling step in more detail.

2.5.6 MILESTONE SELECTION

Item infit and outfit are convenient statistics for selecting the milestones that fit the model. Figure 2.5.5 displays the infit and outfit statistics of model 1339_11. The correlation between infit and outfit is high (r = 0.84). The expected value of the infit and outfit statistics for a perfect fit is 1.0. The centre of infit and outfit in Figure 2.5.5 is approximately 1.0, so on average one could say the items fit the model. Note however that fit values above and below the values of 1.0 are qualitatively different. Item with fit statistics exceeding 1.0 fit the model less well than expected (**underfit**), whereas items with fit statistics lower than 1.0 fit the model better than expected (**overfit**). See Chapter 1, Section 6.1 (van Buuren & Eekhout, 2021) for more details.

Some practitioners remove both underfitting and overfitting items. However, we like to preserve overfitting items and be more strict in removing items that underfit. The idea is that preservation of the best fitting items may increase scale length, and hence reliability and measurement precision. Figure 2.5.5 draws two cut-off lines at 1.0. Taking items with infit < 1.0 and outfit < 1.0 will select **631 out of 1339** items for further modelling.

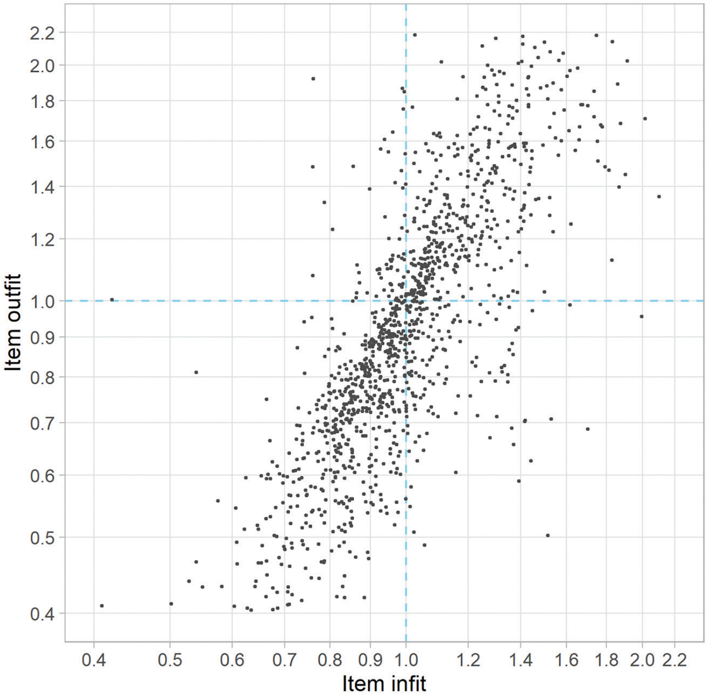


FIGURE 2.5.5 Infit and outfit of 1339 items in model `1339_11`.

About 8 percent of the points falls outside the plot.

A practical problem of item removal is that it also affects equate group composition. By default, a removed item will also be removed from the equate group, so item removal may reduce the size of an equate group below two items. For passive equates this is no problem, since passive equates do no affect the estimates. However, removal of an underfitting item from an active equate group will break the bridge between the instrument it pertains to and the rest of the item set. Potentially this can result in substantial effects on the D-score distribution of the cohort, as demonstrated in Figure 2.5.2. As a solution, we force any items that are members of active equate groups to remain in the analysis. If that leads to substantially worse equate fit in the next model, we must search for alternative equate groups that bridge the same instruments and that are less sensitive to misfit.

2.5.7 OTHER MODELLING ACTIONS

2.5.7.1 Instrument fit

Some instruments fit better than others. Figure 2.5.6 shows the box plots of outfit per instrument. Instruments `bar`, `by1`, `ddi` and `vin` generally fit well,

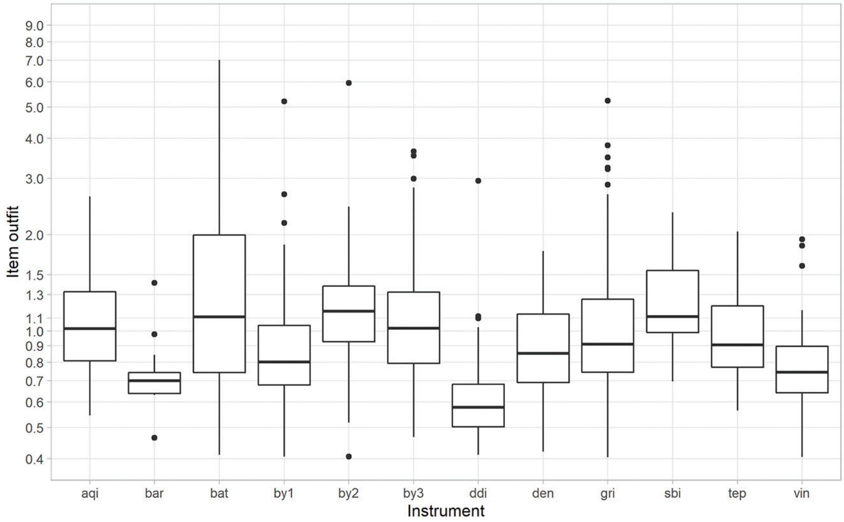


FIGURE 2.5.6 Box plot of the distribution of item outfit per instrument in model `1339_11`.

whereas discrepancies between model and data are larger for `bat`, `by2` and `sbi`. Through additional modelling, we found that it was extremely difficult to get enough high-quality bridge items that could link `bat` (Battelle Development Inventory) to the other instruments. We also found that models without the Battelle were able to better discriminate children in the upper range of the D-score scale. We therefore opted to remove `bat` from the model, even though this meant that one cohort (`GCDG-BRA-2`) had to be dropped from the analysis.

It is not clear why `bat` does not fit. Perhaps the scoring system of the Battelle in three categories invokes scoring behaviour that is different from the PASS/FAIL scoring used by most other instruments, even though this appears to be less of a troublesome aspect in `aqi`, which also uses three response categories.

2.5.7.2 Splitting, combining and selecting equate groups

Most of the modelling effort went into finding a set of high-quality equate groups that link the instruments. For example, we tried to bridge the South-African study placing `vinxxc016` (uses a short sentence) into `EXP26` (two-word sentences) and `EXP36` (sentences of 3 or more words), but neither option led to a reasonable model. On the surface, milestone reasonable model. On the surface, milestone `by3gmd06` (balances on right foot, 2 seconds) appears to fit within `GM60` (balances on foot), but the analysis showed large discrepancies with the other items in the groups, so it had to be taken out.

Subject-matter experts identified 38 items that were thought to be cross-culturally incompatible. Table 2.5.1 provides an overview. Many of such milestones involve a specific language concept (such as a pronoun), refer to stairs (less common

TABLE 2.5.1
Milestones not used for equating because of limited cross-cultural validity.

Item	Label
aqislc023	When you dress your baby does she lift her foot for her shoe, sock, or pant leg?
aqislc041	Using these exact words, ask your child, "Are you a girl or a boy?" Does your child answer correctly?
by1mdd050	Washes and dries hands
by1pdd053	Bowel and bladder control
by1pdd054	manipulates table edge actively
by2pdd069	Walsk up stairs with help
by3cgd043	Walks down stairs with help
by3cgd052	Walks down stairs with help
by3gmd047	Clear Box: Front
by3gmd049	Clear Box: Sides
by3gmd057	Uses pronouns
by3gmd058	Walks Up Stairs Series: Both feet on each step, with support.
by3red030	Walks Down Stairs Series: Both feet on each step, with support
by3exd030	Walks Up Stairs Series: Both feet on each step, alone.
barxxx016	Walks Down Stairs Series: Both feet on each step, alone
barxxx020	Understands pronouns (him, me, my, you, your)
dengmd020	Eats with spoon without help (M; can ask parents)
densld012	Takes off shoes and socks (M; can ask parents)
densld013	Can dress (one piece) (M; can ask parents)
grigmd219	Walk Up Stairs
grigmd222	Drink from a cup
mdsgmd002	help in house
mdsgmd003	(Locomotor) Walks up and down stairs.
mdsgmd004	(Locomotor) Goes alone on the stairs (any method)
mdsgmd005	Hands-and-knees crawling
mdsgmd006	Standing with assistance
ddifmm019	Walking with assistance
ddifmd154	Standing alone
vinxxc002	Walking alone
vinxxc003	chew solid foods
vinxxc009	take off socks / shoes
vinxxc012	get on with other children
vinxxc014	know what's edible
vinxxc022	walk upstairs
vinxxc028	avoid simple danger - knife / hot
vinxxc031	help around the house / clear table
vinxxc040	Play or do things with other children of same age eg sing song
ddifmm025	Help with little things around the house eg pick up things

in rural settings), help in house or clothing behaviour. These items have different meanings in different contexts, so they were not used to bridge instruments.

2.5.8 ITEM INFORMATION

Item information is a psychometric measure that quantifies the sensitivity of the item to changes in the person's ability. An item is most sensitive around the D-score value where the PASS probability equals the FAIL probability, which corresponds to the item difficulty (δ_i). One unit change around δ_i has a large effect on the probability of endorsing, while one unit change far away from δ_i has negligible impact. Suppose person A had passing probability 0.7 for some item. The information delivered by that item for person A is the product 0.7 × (1.0 – 0.7) = 0.21. Suppose person B has a D-score that coincides with the difficulty level of the item. In that case, the information for B equals 0.5 × (1 – 0.5) = 0.25, the maximum. Likewise, for a person C with high ability, the information could be 0.98 x 0.02 = 0.02, so that item carries almost no information for person C.

The information is inversely related to the error of measurement. More information amounts to less measurement error. For each response in the data, we can compute the amount of information it contributed to the model D-score. By summing the information over persons, we obtain a measure of certainty about the difficulty estimate of the item. This sum of information incorporates both the number of administrations and the quality of the match between person abilities and item difficulty.

Figure 2.5.7 displays the summed information for each item, divided into four grades: A(best) to D (worst). The information grade measures the stability of the difficulty estimate. Most items receive grades higher than C. In total, 30 milestones have grade D. Adding these items to future studies may yield important additional information.

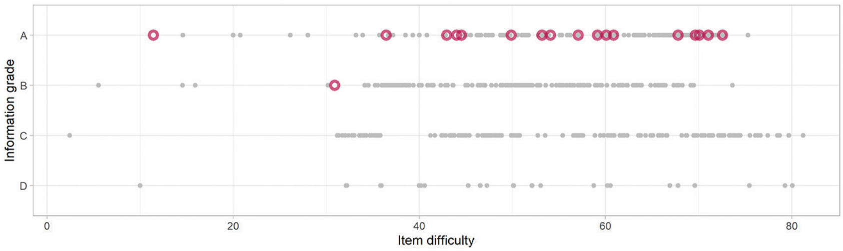


FIGURE 2.5.7 Item information grade by item difficulty for the final model.

TABLE 2.5.2
Equate group information in the final model.

equate	tau	n	info	grade
EXP2	11.44	3608	162.33	A
REC6	30.9	5428	95.40	B
GM25	36.43	6380	470.63	A
FM26	42.93	4155	296.78	A
GM35	44.01	5522	356.04	A
COG36	44.53	7912	230.03	A
GM42	49.86	5953	327.74	A
FM31	53.17	10991	731.66	A
COG55	54.08	5647	420.35	A
FM72	57.07	5430	253.64	A
EXP26	59.15	9119	578.79	A
SA1	60.08	3363	172.11	A
FM38	60.87	10236	491.68	A
FM52	67.8	13487	1159.94	A
FM43	69.66	15765	1563.89	A
GM60	70.09	9519	1070.61	A
REC40	71.04	10393	1182.91	A
FM61	72.56	10612	945.87	A

The red circles indicate active equate groups. Most have grade A, so we have a lot of information about the items that form the active equate groups. Table 2.5.2 displays more detailed information for the active equate groups. The sample sizes are reasonably large. Many information statistics are well is above 100; the criterion for Grade A. The interpretation of this criterion is as follows. Suppose that we obtain a sample of 400 persons who are all perfectly calibrated to the item of interest. In that case, the information for that item will be equal to 100.

2.5.9 FINAL MODEL

Unfortunately, there is no single index of model fit that we can optimize. Modelling is more like a balancing act among multiple competing objectives, such as

- preserving as many items as possible that fit the model;
- finding high-quality active equate groups that span many cohorts and instruments;
- picking active equate groups for which we have enough information;
- providing reasonable age-conditional distributions of the D-score;

- representing various developmental domains in a fair way;
- preserving well-fitting historical models as new data become available;
- maintaining a reasonable calculation time.

This section showed various modelling techniques and ways to assess the validity of the model. In real life, we fitted a total number of 140 models on the data and made many choices that weigh the above objectives. The final model for the GCDG data consists of 565 items (originating from 14 instruments) that fit the Rasch model and that connect through 18 equate groups. Due to the sparseness of data at the very young ages, the quality of the model is best for ages between 4–36 months.

Model `565_18` formed the basis of the publication by Weber *et al.* (2019). Additional detail on model `565_18` is available through the `dmodel` shiny app at https://tnochildhealthstatistics.shinyapps.io/dmodel/.