2.3 Comparability

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 describes challenges and methodologies to harmonize child development measurements obtained by different instruments:

- Are instruments connected? (2.3.1)
- Bridging instruments by mapping items (2.3.2)
- Overview of promising item mappings (2.3.3)

2.3.1 ARE INSTRUMENTS CONNECTED?

The ultimate goal is to compare child development across populations and cultures. A complication is that measurements are made by different instruments. To do deal with this issue, we harmonize the data included in the GCDG cohorts. In particular, we process the milestone responses such that the following requirements hold:

- Every milestone in an instrument has a unique name and a descriptive label;
- Every milestone occupies one column in the dataset;
- Item scores are (re)coded as: 1 = PASS; 0 = FAIL;
- Items not administered or not answered are a missing value;
- Every row in the dataset corresponds to a unique cohort-child-age combination.
- Two cohorts are indirectly connected if both connect to a third cohort that connects them.
- Two instruments are indirectly connected if both connect to a third instrument that connects them.
- Any differences between studies can be attributed to the difficulties of the instruments.

Cohorts and milestones need to be *connected*. There are several ways to connect cohorts:

- Two cohorts are directly connected if they use the same instrument;

DOI: 10.1201/9781003216315-14

- Two cohorts are indirectly connected if both connect to a third cohort that connects them.
- Two instruments are indirectly connected if both connect to a third instrument that connects them.
- Any differences between studies can be attributed to the difficulties of the instruments.

Likewise, instruments can be connected:

- Two instruments are directly connected if the same cohort measures both;
- Two instruments are indirectly connected if both connect to a third instrument that connects them.
- Any differences between studies can be attributed to the difficulties of the instruments.

An `X` in Table 2.3.1 identifies which cohorts use which instruments. The linkage table shows that studies from China, Colombia, and Ethiopia are directly connected (by `by3`). Brazil 1 indirectly connects to these studies through `den`. Some cohorts (e.g., Chile 1 and Ecuador) do not link to any other study. Likewise, we might say that `aqi`, `bat`, `by3`, and `den` are directly connected. Note that no indirect connections exist to this instrument group.

Table 2.3.1 is a somewhat simplified version of the linkage pattern. As we saw in section 2.2.2, there are substantial age differences between the cohorts. The linked instrument linkage table shows the counts of the number of registered scores per age group. What appears in Table 2.3.1 as one test may consist of two disjoint subsets, and hence some cohorts may not be connected after all.

Connectedness is a necessary - though not sufficient - requirement for parameter identification. If two cohorts are not connected, we cannot distinguish between the following two alternative explanations:

- Any differences between studies can be attributed to the ability of the children;
- Any differences between studies can be attributed to the difficulties of the instruments.

The data do not contain the necessary information to discriminate between these two explanations. Since many cohorts in Table 2.3.1 are unconnected, it seems that we are stuck.

The next section suggests a way out of the dilemma.

2.3.2 BRIDGING INSTRUMENTS BY MAPPING ITEMS

Many instruments for measuring child development have appeared since the works of Shirley (1933) and Gesell (1943). It is no surprise that their contents

TABLE 2.3.1
Linkage pattern indicating combinations of cohorts and instruments.

	aqi	bar	bat	by1	by2	by3	ddi	den	gri	mac	peg	sbi	sgr	tep	vin
Bangladesh					x										
Brazil 1								x							
Brazil 2			x												
Chile 1				x											
Chile 2			x											x	
China						x									
Colombia 1						x									
Colombia 2	x		x			x		x							
Ecuador		x													
Ethiopia						x									
Jamaica 1									x						
Jamaica 2									x						
Madagascar										x	x	x			
Netherlands1							x								
Netherlands2							x								
South Africa					x				x				x		x

show substantial overlap. All tools assess events like starting to see, hear, smile, fetch, crawl, walk, speak, and think. We will exploit this overlap to bridge different instruments. For example, Table 2.3.2 displays the labels of milestones from six instruments. All items probe the ability of the child to formulate "sentences" of two words.

The idea is to check whether these milestones measure development in the same way. If this is found to be true, then we may formally restrict the

TABLE 2.3.2
Example of similar items from different instruments.

Item	Label
by1mdd136	Sentence of 2 words
by2mdd114	Uses a two-word utterance
ddicmm041	Says sentences with 2 words
denlgd019	Combine Words
grihsd217	Uses word combinations
vinxxc016	Use a short sentence

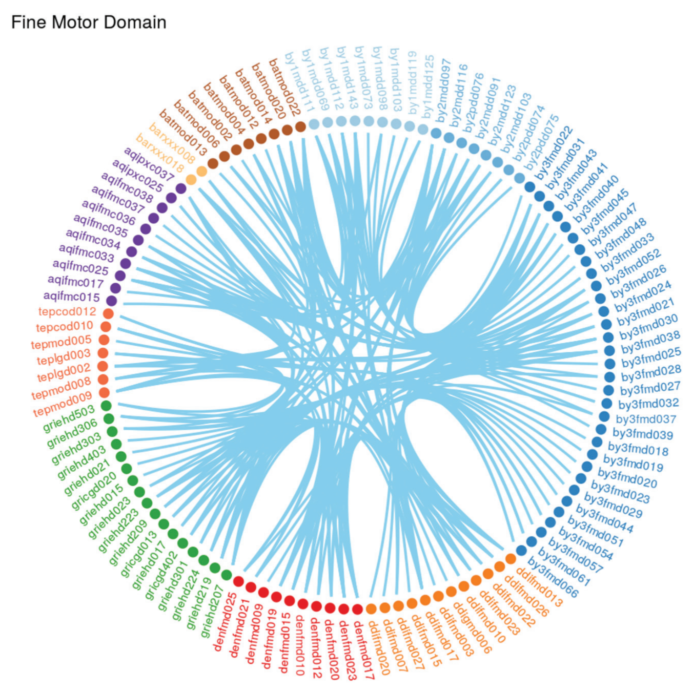

FIGURE 2.3.1 Connections between the instruments via mapped item groups for the fine motor domain (https://tnochildhealthstatistics.shinyapps.io/GCDG_mapping/).

difficulty levels of these milestones to be identical. This restriction provides a formal bridge between the instruments. We repeat the process for all groups of similar-looking items.

A first step in the bridging process is to group items from different instruments by similarity. As the `by3` is relatively long and is the most often used instrument, it provides a convenient starting point. Subject matter experts experienced in child development mapped items from other tools to `by3` items. These experts evaluated the similarity of wordings and descriptions in reference manuals. Also, they mapped same-skill items across other instruments into groups if these did not map onto `by3` items.

Figure 2.3.1 connects similar items and hence visualizes connections between instruments for the fine motor domain. Items are displayed in the wheel, coloured by instrument. In the online application we organized item mappings into five domains: fine motor (`FM`), gross motor (`GM`), cognitive (`COG`), receptive (`REC`), and expressive (`EXP`). The `Prev` and `Next` buttons allow us to visit other domains.

2.3.3 AGE PROFILE OF ITEM MAPPINGS

Another way to explore the similarity of milestones from different instruments is to plot the probability of passing by age. Figure 2.3.2 shows two examples. The first graph presents the age curves of a group of four cognitive items for assessing the ability to put a cube or block in a cup or box. The milestones are administered in different studies and seem to work similarly. The second plot shows a similar graph for items that assess the ability to build a tower of six cubes or blocks. These milestones have similar age patterns as well.

Figure 2.3.3 presents two examples of weak item mappings. Notable timing differences exist for the “babbles” and “bangs” milestones, which suggests that we should not take these as bridges.

While these plots are suggestive, their interpretation is surprisingly complicated. We may find that age profiles of two milestones *A and B* administered in samples 1 and 2, respectively, *are identical* if

- A and B are equally difficult and samples 1 and 2 have the same maturation level;
- A is more difficult than B and sample 1 is more advanced than 2.

Similarly, we may find that the age profile for *A is earlier than B* if

- A is easier than B and if samples 1 and 2 have the same level of maturation;
- A and B are equally difficult and if sample 1 is more advanced than sample 2.

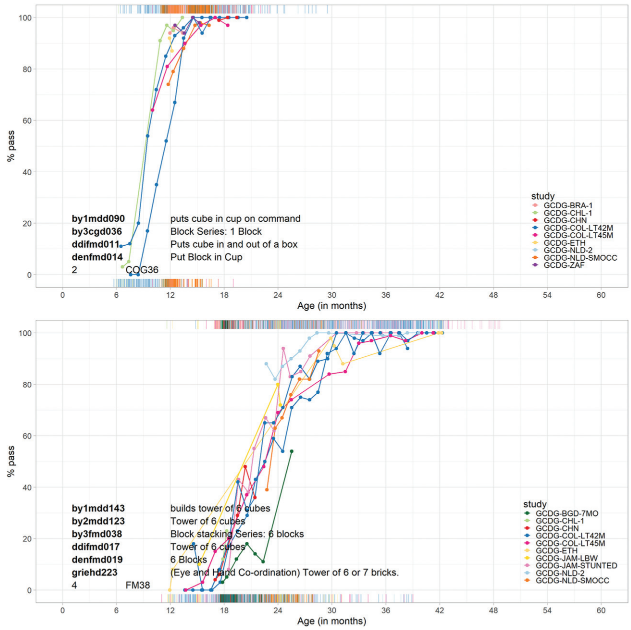

FIGURE 2.3.2 The probability of passing by age in potential bridging items.

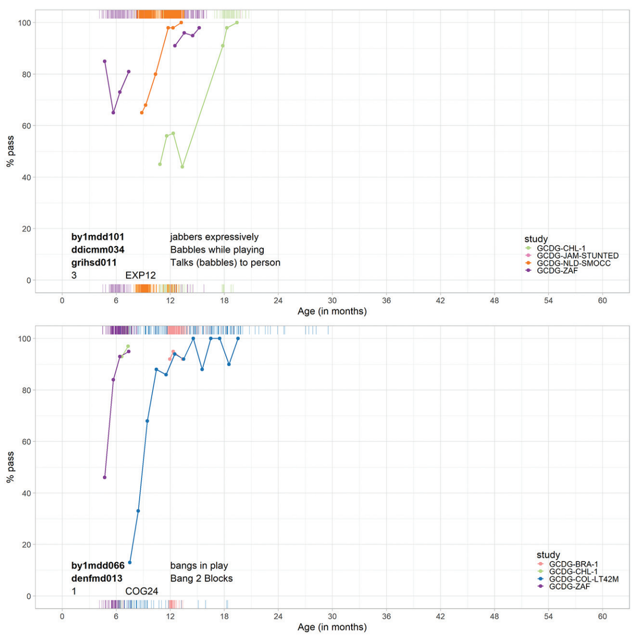

FIGURE 2.3.3 Probability to pass items for age in poor bridges.

Note that the age curves confound difficulty and ability, and hence cannot be used to evaluate the quality of the item map.

What we need to do is separate difficulty and ability. For this, we need a formal statistical model. The next section introduces the concepts required in such a model.