

1.7 Validity

Stef van Buuren^{1,2}

Iris Eekhout¹

¹Netherlands Organisation for Applied Scientific Research TNO, Leiden, 2316 ZL, The Netherlands

²University of Utrecht, Utrecht, 3584 CH, The Netherlands

Validity is a generic term that refers to the question of how well an instrument measures what it claims to measure. There are various aspects of validity. This section briefly reviews the main types of validity:

- Internal validity (1.7.1)
- External validity (1.7.2)

1.7.1 INTERNAL VALIDITY

1.7.1.1 CONTENT VALIDITY

Content validity is the extent to which the D-score represents all facets of development. In contrast to “face validity,” which assesses whether the test appears valid to respondents, content validity is about what is measured.

One important form of content validity is that we wish to make sure that the measurement scale represents the various developmental domains in a fair way. In the simplest case, we can assign each milestone uniquely to one domain and evaluate coverage by splitting the cumulative item information.

Figure 1.7.1 shows the coverage of the three domains of the DDI at various levels of the D-score. The three domains of the DDI are relevant at most ability

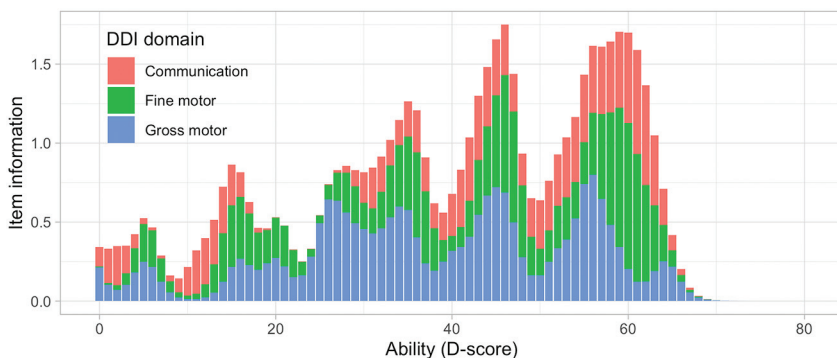


FIGURE 1.7.1 Cumulative item information by DDI domain.

levels. The DDI contains no communication milestones between 20 *D* and 30 *D*, so at these levels, the DDI measures primarily motor performance.

Content validity assessment is part of modelling when we examine what milestones fit the model. Content validity also means that all relevant facets of development are measured. As discussed in Section 1.6.1, we may remove items that do not fit the model and hence fail to measure development in the technical sense. As a result, we may lose items considered relevant by subject-matter specialists. If we want to preserve these, we could fit a separate model that captures another development aspect. We did not encounter the issue with the DDI. In contrast, our finding that items allocated to different domains form a unidimensional scale underlines the content validity of the D-score.

1.7.1.2 CONSTRUCT VALIDITY

Construct validity is the extent to which the D-score behaves like the theory says the construct should behave. For example, we expect that child development advances with age. Figure 1.4.3 provides convincing evidence that the D-score increases fastest in the first six months and keeps rising at a slower rate as children age. This phenomenon is consistent with theories in growth and child development.

In Section 1.4, we assumed that child development is a latent variable. Figure 1.7.2 provides one way to evaluate the validity of this assumption. The figure plots the item fit for each milestone coloured by domain. Items from different domains fit equally well, so there is no evidence that the D-score favours a particular area. Put in more technical terms; the DDI domains do not explain differences in the item fit residuals of the model.

1.7.2 EXTERNAL VALIDITY

1.7.2.1 DISCRIMINATORY VALIDITY

Discriminatory validity indicates the extent to which the D-score can distinguish children with non-normal development from children that are developing normally. We may evaluate this by identifying children with lagging development, for example, indicated by reflex or tonus problems, and study whether the D-score can discriminate those children from the general population. Section 1.9.3 presents some examples.

1.7.2.2 CONVERGENT AND DIVERGENT VALIDITY

Convergent validity is the extent to which the D-score relates to similar constructs. We measure it by the correlation between the D-score and the total score on Bayley-III or Denver.

The correlation with the other construct should be 0.6, or higher for good convergent validity. Unfortunately, at present, only limited data is available for the DDI, so we cannot assess convergent validity for the D-score at this point.

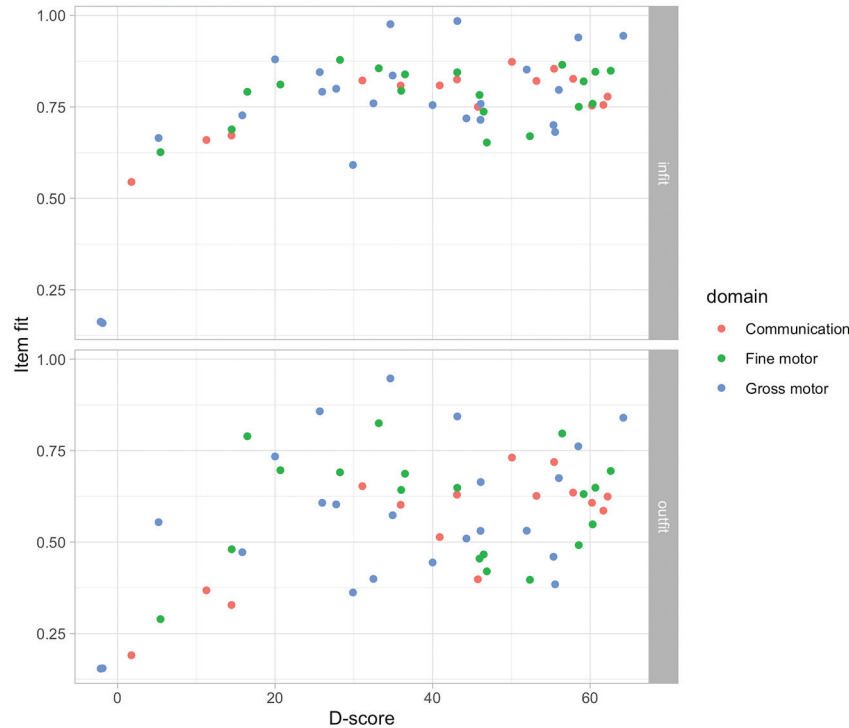


FIGURE 1.7.2 Item fit by D-score for the DDI domains.

Divergent validity is the extent to the D-score is uncorrelated with measures of a different construct.

Figure 1.7.3 shows both convergent and divergent validity at work. The figure shows that, as expected, there is a strong and almost linear relation between body height and the D-score. However, after correction for the child’s age, the relationship between height and D-score almost disappears. Thus, growth and development are entirely different concepts.

We can also evaluate the strength of the relations between the D-score and proxy measures of child development, such as stunted height growth (see section 1.1.3). The low correlation between DAZ and HAZ suggests that stunting is a poor proxy for child development.

1.7.2.3 PREDICTIVE VALIDITY

Predictive validity refers to the degree to which the D-score predicts the score on a criterion that is measured later. For the D-score, we may compare to measures for IQ at the school-age as a possible criterion.

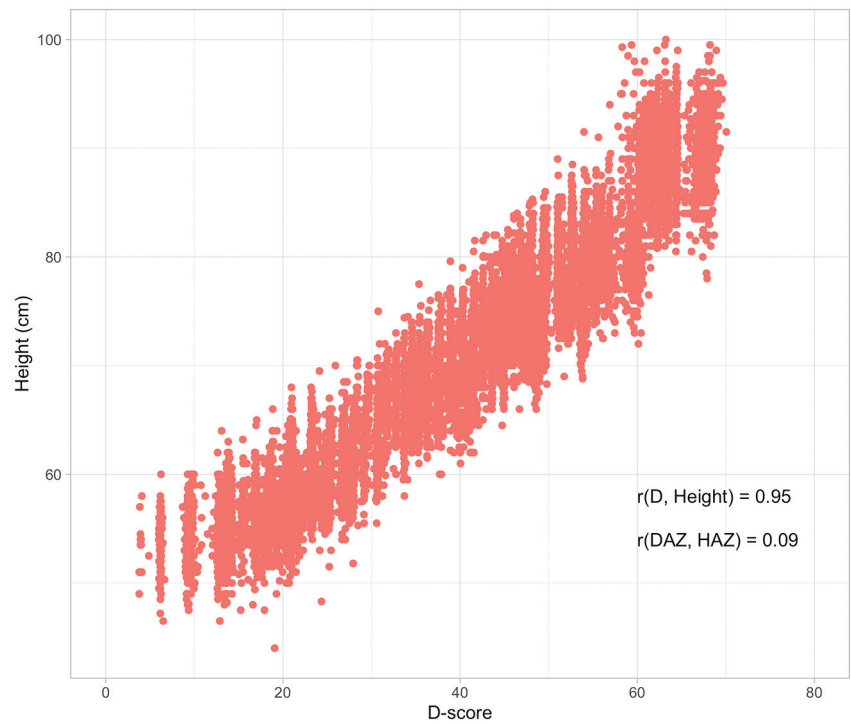


FIGURE 1.7.3 Relation between body height and the D-score in the SMOCC data.

Vlasblom *et al.* (2019) found strong evidence that individual milestones of the DDI measured during the first years of life predict later intellectual functioning at ages 5–10 years. It is to be expected that the D-score, which builds upon these individual items, will also predict limited intellectual functioning, perhaps even better.