



TwinLife

TwinLife Technical Report Series

No. 02, April 2017

# Documentation TwinLife Data: Cognitive Abilities

by Juliana Gottschling

[j.gottschling@mx.uni-saarland.de](mailto:j.gottschling@mx.uni-saarland.de)





# TwinLife

**Juliana Gottschling**

**Documentation TwinLife Data: Cognitive Abilities**

TwinLife Technical Report Series No. 02

Project TwinLife “Genetic and social causes of life chances”

Bielefeld, April 2017

TwinLife Technical Report Series

General Editors: Martin Diewald, Rainer Riemann and Frank M. Spinath

ISSN 2512-403X

This publication has been funded by the German Research Foundation (DFG).

TwinLife Technical Reports are refereed scholarly papers. Submissions are reviewed by the general editors before a final decision on publication is made.

The Technical Report Series is a forum for presenting technical works (e.g., data documentation, field reports etc.) in progress. Readers should communicate comments on the manuscript directly to the author(s).

The papers can be downloaded from the project website:

<http://www.twin-life.de/en/twinlife-technical-report-series>

TwinLife “Genetic and social causes of life chances”

University of Bielefeld

Faculty of Sociology

PO Box 100131

D-33501 Bielefeld

Germany

Phone: +49 (0)521 106-4309

Email: [martin.diewald@uni-bielefeld.de](mailto:martin.diewald@uni-bielefeld.de)

Web: <http://www.twin-life.de/en>



<b>Content</b>	<b>Page</b>
1. Description of the cognitive test battery	2
2. Calculation of the cognitive test scores	3
3. Handling of invalid cases	4
4. Recommendation for the use of the cognitive test	6
 <b>References</b>	 6
 <b>Appendix</b>	
Figure 1: Flow diagram of invalid case decision	7
Table 1: Total concordance rate per test	8
Table 2: Number of (sub-)tests set to NA	8
Table 3: List of reasons for exclusion	9

## 1. Description of the cognitive test battery

We used three (children, aged 9 and younger), respectively four (adults, aged 10 and older) subtests of the CFT (Culture Fair Test; German: Grundintelligenztestskala), a widely used and well validated cognitive test battery that captures non-verbal (fluid) intelligence as a proxy for general cognitive ability.

Depending on the age, the cognitive test was administered either as a paper-pencil test by trained interviewers (for children), or as a computer version of the test (for adults). The PC version was designed in accordance to the validated online version of the CFT<sup>1</sup>.

### Children, aged 9 and younger

- Test: CFT 1-R (Grundintelligenztestskala 1, Revision; Weiß & Osterland 2012)
- Norm sample: 5;3 to 9;11 years of age
- Three subtests (item examples):

- Subtest 1: Figural Reasoning – 15 items



- Subtest 2: Figural Classification – 15 items



- Subtest 3: Matrices – 15 items



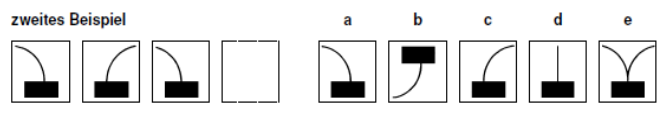
- Paper pencil administration by trained interviewer; Test instruction was in accordance to the guidelines described in the CFT 1-R manual. Interviewers were provided with a standardized test instruction; the given test time (see CFT 1-R manual for further information) was timed by a programmed test module; All children of a family were tested at the same time whenever possible. Interviewers had the possibility to comment on the test situation and the behavior of the children.
- Test time: approx. total test time of 30 minutes including test instruction. Each test was administered in a short version (3 minutes) and a long version (+1 minute), in accordance with the CFT 1-R test instructions. The kids were allowed to correct their answers in the additional minute (happened only occasionally).

<sup>1</sup> The use and technical implementation of the CFT subscales was reviewed and approved by Hogrefe

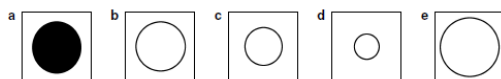
## Children and adults, aged 10 and older

- Test: CFT 20-R (Grundintelligenztestskala 2, Revision; Weiß 2006)
- Norm sample: 8;5 to 60 years of age
- Four subtests (item examples):

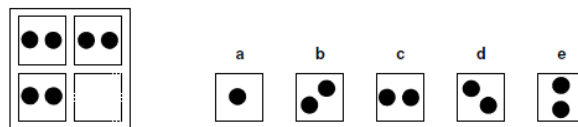
- Subtest 1: Figural Reasoning – 15 items



- Subtest 2: Figural Classification – 15 items



- Subtest 3: Matrices – 15 items



- Subtest 4: Reasoning – 11 items



- Test administered as a PC version in accordance with the original computer version of the test; Instructions were given on the screen with a standardized test time as given in the manual.
- Test time: each test was administered in a short (subtests 1 and 2: 4 minutes; subtests 3 and 4: 3 minutes) and long version (+ 1 minute each); The additional test minute started automatically if the test subject did not finish the test in the short time. It was allowed to skip items; Skipped items were presented again at the end of the test if the test time was not expired. It was also possible to go back to specific items and to modify the answer (which had consequences for the calculation of the total test score, see [2]).

## 2. Calculation of the cognitive test scores

Each correct answer on a test item was coded with 1, each wrong answer was coded with 0. Item non-responses were coded with -88. Items worked on in the short test version were coded as -89 in the long test version. Multiple item responses in the paper and pencil test for children were coded as -87. Technical recording errors in the computer assisted test for adults were coded as -94.

The sumscores for each subtest are computed as sum of all correctly solved items; For each subtest, three different scores were calculated: 'short' (sum of all correct items in the short test time), 'long' (additionally given answers), and 'total' (sum of all correctly solved items, independent of test time, last given answer was counted in case of answer modification).

For CFT 20-R: As participants were allowed to modify their answers, the 'long' score can also take negative values if participants changed correct answers into wrong answers. A value of Zero does not necessarily imply that there were no changes, it could also mean that the participant made some changes in the wrong direction and the same number of changes in the right direction.

This procedure results in 9 subscores for children, and 12 for adults. Variables are labelled as follows:

- igf0180, igf0181, igf0182 : CFT 20-R Test1 short: sum score; long: sum score; total: sum score
- igf0280, igf0281, igf0282 : CFT 20-R Test2 short: sum score; long: sum score; total: sum score
- igf0380, igf0381, igf0382 : CFT 20-R Test3 short: sum score; long: sum score; total: sum score
- igf0480, igf0481, igf0482 : CFT 20-R Test4 short: sum score; long: sum score; total: sum score
- igf0580, igf0581, igf0582 : CFT 1-R Test1 short: sum score; long: sum score; total: sum score
- igf0680, igf0681, igf0682 : CFT 1-R Test2 short: sum score; long: sum score; total: sum score
- igf0780, igf0781, igf0782 : CFT 1-R Test3 short: sum score; long: sum score; total: sum score

### **3. Handling of invalid cases**

Because invalid cognitive test scores (i.e., scores that result from technical errors, test situation, or other reasons, rather than the true ability of a participant) can distort the results of statistical analyses, we implemented a structured control mechanism in order to identify cases with conspicuous test scores to minimize invalid cognitive test data. Invalid test scores can occur if, for example, the participant answered at random, or with a response pattern (e.g., sole use of one response option). Also, cases in which the test score (or the number of worked items) of one subtest extremely differ from the remaining subtests, or in which the interviewer noted difficulties in the test situation, were treated as "invalid" (see Figure 1 Appendix for a complete listing of peculiarities that were rated as "potentially invalid").

Final (sub-)test scores of cases that were rated as “invalid” were set to NA, i.e. were coded as -86 in the dataset. However, we kept the original coding of the single items as wrong/right and did not set the single items to -86 in these cases.

- In total, 8,079 individuals took part in the cognitive test in Wave1.1;
  - 6,877 adults, and 1,202 children in Wave1.1
- In a first step, test scores of children younger than 5 years of age by the time of testing were set to -86, given that the CFT 1-R is not applicable in this age group.
  - 51 cases in Wave1.1
- Cases with a reasonable score (i.e., higher score than would be expected by chance) and enough answered items (based on the average number of answered items for each test within the population) were defined as OKAY
  - 6,400 adults, and 980 children in Wave1.1
- Participants with a lower score than would be expected by chance and/or only few answered items (based on the average number of answered items for each test within the population) were identified as potentially invalid cases which need to be controlled.
  - 478 adults and 171 children as test cases in Wave1.1
- In a first control step, we identified all cases with missings in all subtests, i.e., participants that did not work at a single item. This also include cases in which the computer program may not have worked correctly and did not save the given answers.
  - 169 adults, and 20 children in Wave1.1
- The remaining cases were evaluated according to the described decision tree displayed in Appendix (Flow diagram) by two independent raters. All cases, in which the two raters did not come to the same conclusion regarding the exclusion of the (sub-)test scores of a participant, were additionally rated by an independent third rater. This procedure led to high concordance rates of over 80% between the first two raters (see Table 1 for details).

#### Detailed description of the rating process:

Table 1 (see Appendix) displays the overall concordance rate between the two raters, and the number of unequal ratings per test. Table 2 (see Appendix) contains the number of replaced (with -86) test scores for each test separately. Table 2 also gives the percentage of replaced data in relation to the available test data (only children older than 4 included). The percentage of replaced data is higher for kids, as the raters had more information about the test situation derived from the interviewer comments. For adults, only the test time, the number of worked items per test, and the evaluation of potential response patterns could be used for the decision, and it has been decided to use a conservative approach (i.e., to not set too many cases to -86). Table 3 (see Appendix) lists the reasons based on which the subtest scores were set to -86.

#### 4. Recommendation for the use of the cognitive test

We strongly recommend using the sum scores of the subtests as provided in the SUF. These sum scores are corrected (i.e., (sub-)test scores set to -86) for invalid cases, identified by the procedure described above. However, if users wish to use the uncorrected scores, they can still calculate them by calculating a new sum score, as the answers (right/wrong) are still original and not corrected. Users can either use the `_short` score (i.e. short test time) or the `_total` score (i.e. last given answers independent of test time) of the three, respectively four subtests.

When using the child's scores, users need to be aware that there are cases with children younger than 5. We recommend to not include them in the analyses, as the CFT 1-R is not normed for this age group (in the provided sum scores, the scores were already set to -86 for those cases).

We also recommend to use a latent factor approach whenever possible, and to use an adequate algorithm to substitute missing data. When substituting missing values, we recommend excluding those cases that did not take part in the cognitive test prior to any procedure.

As the data set includes twins from four different age cohorts, we recommend conducting all analyses within cohorts. For analyses across cohorts, the use of adequate procedures to control for age effects is mandatory. This recommendation also applies to the cognitive test data from the twins' parents. For comparisons with the normative sample of the CFT, see Weiß (2006) as well as Weiß and Osterland (2012).

#### References

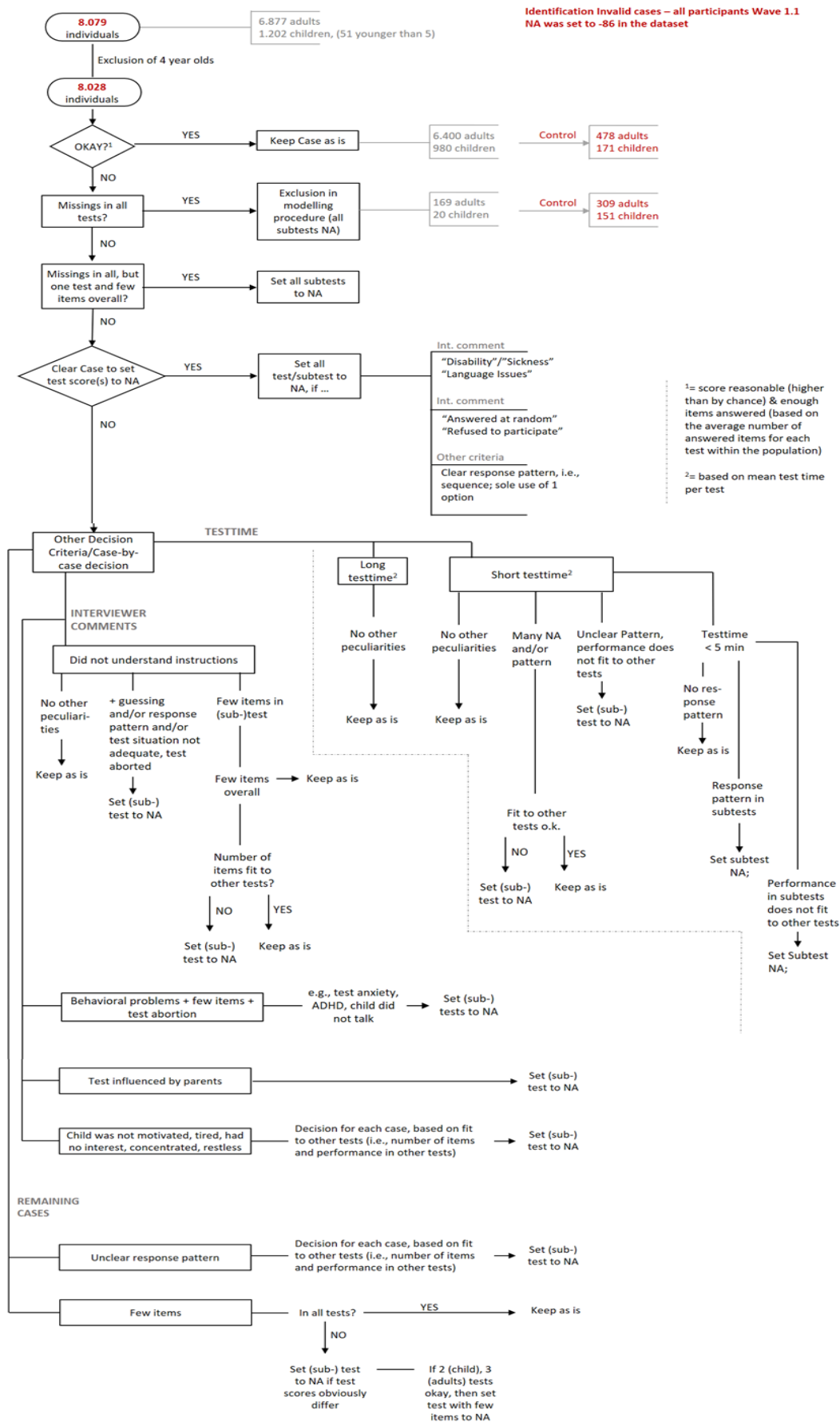
Weiß, R. (2006): CFT 20-R. Grundintelligenztest Skala 2. Manual. Göttingen: Hogrefe.

Weiß, R. H., & Osterland, J. (2012). Grundintelligenztest Skala 1 - Revision: CFT 1-R. Göttingen: Hogrefe.



# Appendix

## Figure 1: Flow Diagram Wave1.1



**Table 1:** Total concordance rate per test

Wave1.1	Children	Adults
No of rated cases	151	309
No of unequal ratings		
Test 1	10	14
Test 2	6	9
Test 3	8	9
Test 4	./.	11
<b>Total</b>	24	43
<b>% concordance</b>	<b>83.8%</b>	<b>86%</b>

**Table 2:** Number of (sub-)tests set to NA<sup>1</sup>

Wave1.1	Children	Adults
	No set to NA	No set to NA
No of rated cases <sup>2</sup>	151	309
Test 1	37	45
Test 2	26	38
Test 3	29	24
Test 4	./.	60
<b>Total</b>	92	167
<b>% replaced data in rated cases</b>	<b>60.9%</b>	<b>54.4%</b>
No cases with all test scores set NA	11	7
<b>% replaced data in total sample (age&gt;4; all NA cases counted once)</b>	<b>5.7%</b>	<b>2.2%</b>

<sup>1</sup> Values were set to -86 in the dataset<sup>2</sup> Does not include cases with missings in all tests and missings in all tests but one

**Table 3:** List of reasons for exclusion

Wave1.1		Test 1	Test 2	Test 3	Test 4
<b>Children</b>	<b>n</b>	<b>37</b>	<b>26</b>	<b>29</b>	
Response pattern [PA]		5	6	4	./.
Short test time [T]		2	2		./.
Few items compared to other tests [FI]		10	1		./.
Performance did not fit to other tests [PF]				3	./.
Case overall conspicuous		13	8	9	./.
<i>Interviewer Comments [IC]</i>					./.
[1] Disability/Sickness/Language Issues			1	4	./.
[2] Answered at random/Refused to participate			3 (+ PA) 2 (+ FI)		./.
[3] Did not understand instructions		4 (+ FI) 1 (+ IC 5)		3 (+ FI) 1 (+ PA)	./.
[4] Behavioral problems + few items + test abortion		1 (+ FI)		1 (+ IC 6)	./.
[5] Test influenced by parents					./.
[6] Child not motivated, tired, no interest, not concentrated, restless		1 (+ FI)	3 (+ FI)	1 (+ PA) 3 (+ FI)	./.
<b>Adults</b>	<b>n</b>	<b>45</b>	<b>38</b>	<b>24</b>	<b>60</b>
Response pattern		15	18	16	10
Short test time + other		5	2	3	10
Long test time + other		3	1		1
Few items compared to other tests		11	8	4	11
Performance did not fit to other tests		7	2	1	12
Case overall conspicuous		4	7		16