



universität  
wien

# DIPLOMARBEIT

Titel der Diplomarbeit

„Testtheoretische versus inhaltliche Itemselektion  
anhand des AID 2 (Version 2.1 sowie Version 2.2) -  
Eine retrospektive Analyse“

Verfasser

Lorenz Karl Urban

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, 2014

Studienkennzahl lt. Studienblatt:

A 298

Studienrichtung lt. Studienblatt:

Diplomstudium Psychologie

Betreuer:

Univ.-Prof. i. R. Mag. Dr. Klaus Kubinger



## **Vorwort**

Die Diplomarbeitsphase war von vielen Höhen und Tiefen geprägt. Die Tiefen waren rückblickend vor allem einer mangelnden Überzeugung dem eigenen Schaffen und Tun gegenüber zuzuschreiben.

Für die besonders intensive und lehrreiche Unterstützung bei der Planung des ersten Diplomarbeitsthemas, welches kurz vor Beginn der Erhebung der Stichprobe ad acta gelegt werden musste, möchte ich mich herzlich bei Frau MMag. Lisbeth Weitensfelder bedanken. Auch für Ihren Einsatz und dem von Ihr vermittelten Wissen, im damals noch abgehaltenen Praktikumsseminar "Psychologisches Diagnostizieren", möchte ich Ihr gegenüber meine besondere Wertschätzung zum Ausdruck bringen. Das Praktikumsseminar, welches leider nicht mehr Teil des Psychologiestudiums ist, war für mich wegweisend.

Ein beständiger Begleiter bei beiden Diplomarbeitsthemen war Herr Mag. Jan Steinfeld, dem ich für seine Unterstützung bei der Realisierung der Abschlussarbeit danken möchte.

Die vorliegende Diplomarbeit wäre ohne die Unterstützung von Frau Dr. Gerlinde Kaufmann und Frau Dr. Ilse Aumüller nicht möglich gewesen. Ihnen möchte ich meinen besonderen Dank aussprechen. Aufgrund der von Dr. Kaufmann und Dr. Aumüller zur Verfügung gestellten Testprotokolle wurde diese Arbeit erst ermöglicht. Durch beider herzlichen Umgang mit mir, konnten auch schwierigen Phasen gemeistert werden.

Schließlich gilt mein Dank Professor Dr. Klaus Kubinger, der mich immer wieder auf den richtigen Weg brachte und mich nicht nur im Forschungsseminar für Diplomanden mit Rat und Tat zu unterstützen wusste. Vor allem sei dafür gedankt, dass nach dem Wegfall des ersten Themas rasch ein Neues gefunden wurde, bei dessen Ausgestaltung die bereits für das ursprüngliche Thema geleistete Arbeitszeit mit berücksichtigt wurde.



# Inhaltsverzeichnis

<b>I. EINLEITUNG</b>	<b>1</b>
<b>II. THEORETISCHER TEIL</b>	<b>2</b>
<b>1. Vorteile eines Tests, der dem dichotomen logistischen Modell von Rasch entspricht</b>	<b>2</b>
<b>2. Das Adaptive Intelligenzdiagnostikum (AID)</b>	<b>4</b>
2.1 Aufbau des AID	5
<b>III. EMPIRISCHER TEIL</b>	<b>8</b>
<b>3. Ziel der Untersuchung</b>	<b>8</b>
<b>4. Stichprobe</b>	<b>8</b>
4.1.1 Alter	9
4.1.2 Muttersprache	10
4.1.3 Schulform	10
4.1.4 Durchschnittlich erzielte Leistungen	11
<b>5. Ergebnisse</b>	<b>12</b>
5.1 Untertest 1 Alltagswissen	15
5.2 Untertest 2 Realitätssicherheit	28
5.3 Untertest 3 Angewandtes Rechnen	33
5.4 Untertest 4 Soziale und Sachliche Folgerichtigkeit	39
5.5 Untertest 6 Synonyme Finden	40
5.6 Untertest 9 Funktionen Abstrahieren	50
5.7 Untertest 10 Analysieren und Synthetisieren-abstrakt	57
5.8 Untertest 11 Soziales Erfassen und Sachliches Reflektieren	58
<b>6. Diskussion</b>	<b>65</b>
6.1 Itemanalyse für den UT 1	66
6.2 Itemanalyse für den UT 2	70
6.3 Itemanalyse für den UT 3	70
6.4 Itemanalyse für den UT 4	70
6.5 Itemanalyse für den UT 6	71

6.6	Itemanalyse für den UT 9	71
6.7	Itemanalyse für den UT 10	72
6.8	Itemanalyse für den UT 11	72
6.9	Synopsis	75
<b>7.</b>	<b>Zusammenfassung</b>	<b>78</b>
<b>IV.</b>	<b>LITERATURVERZEICHNIS</b>	<b>81</b>
<b>V.</b>	<b>ANHANG</b>	<b>83</b>
<b>8.</b>	<b>Abstract - deutsch</b>	<b>83</b>
<b>9.</b>	<b>Abstract - english</b>	<b>84</b>
<b>10.</b>	<b>Curriculum Vitae</b>	<b>85</b>

## Abbilungsverzeichnis

<b>Abbildung 5.1:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 vor dem Ausschluss von Aufgaben.	16
<b>Abbildung 5.2:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 1 vor dem Ausschluss von Aufgaben.	16
<b>Abbildung 5.3:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 vor dem Ausschluss von Aufgaben; die Ergebnisse werden nur unter Vorbehalt dargestellt.	16
<b>Abbildung 5.4:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i13.	17
<b>Abbildung 5.5:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i20.	18
<b>Abbildung 5.6:</b> Grafische Modellkontrolle des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i23.	18
<b>Abbildung 5.7:</b> Grafische Modellkontrolle des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i31.	19
<b>Abbildung 5.8:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i42.	19
<b>Abbildung 5.9:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i57.	20
<b>Abbildung 5.10:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 1 mit Standardschätzfehler des Items i12.	20
<b>Abbildung 5.11:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 1 mit Standardschätzfehler des Items i26.	21
<b>Abbildung 5.12:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i16.	22
<b>Abbildung 5.13:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i48.	22
<b>Abbildung 5.14:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i50.	23
<b>Abbildung 5.15:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i51.	23
<b>Abbildung 5.16:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i54.	24
<b>Abbildung 5.17:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i55.	24
<b>Abbildung 5.18:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 1 nach dem Ausscheiden von vierzehn Items.	27
<b>Abbildung 5.19:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 1 nach dem Ausscheiden von vierzehn Items.	28

<b>Abbildung 5.20:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 1 nach dem Ausscheiden von vierzehn Items.	28
<b>Abbildung 5.21:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 2 vor dem Ausschluss von Aufgaben.	29
<b>Abbildung 5.22:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 2 vor dem Ausschluss von Aufgaben.	29
<b>Abbildung 5.23:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 2 vor dem Ausschluss von Aufgaben.	30
<b>Abbildung 5.24:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 2 mit Standardschätzfehler des Items i8.	30
<b>Abbildung 5.25:</b> Graphische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 2 nach dem Ausscheiden von einem Item.	32
<b>Abbildung 5.26:</b> Graphische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 2 nach dem Ausscheiden von einem Item.	32
<b>Abbildung 5.27:</b> Graphische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 2 nach dem Ausscheiden von einem Item.	33
<b>Abbildung 5.28:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 3 vor dem Ausschluss von Aufgaben.	34
<b>Abbildung 5.29:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 3 vor dem Ausschluss von Aufgaben.	34
<b>Abbildung 5.30:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 11 vor dem Ausschluss von Aufgaben.	35
<b>Abbildung 5.31:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 3 mit Standardschätzfehler des Items i23.	36
<b>Abbildung 5.32:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 3 mit Standardschätzfehler des Items i24.	36
<b>Abbildung 5.33:</b> Graphische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 3 nach dem Ausscheiden von zwei Items.	37
<b>Abbildung 5.34:</b> Graphische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 3 nach dem Ausscheiden von zwei Items.	37
<b>Abbildung 5.35:</b> Graphische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 3 nach dem Ausscheiden von zwei Items.	38
<b>Abbildung 5.36:</b> Grafische Modellkontrolle des Untertests 4 mittels des Teilungskriteriums Score im UT 2.	40
<b>Abbildung 5.37:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 vor dem Ausschluss von Aufgaben.	41
<b>Abbildung 5.38:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 6 vor dem Ausschluss von Aufgaben.	42

<b>Abbildung 5.39:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 6 vor dem Ausschluss von Aufgaben; die Ergebnisse werden nur unter Vorbehalt dargestellt.	42
<b>Abbildung 5.40:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i48.	43
<b>Abbildung 5.41:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i30.	44
<b>Abbildung 5.42:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i4.	44
<b>Abbildung 5.43:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i5.	45
<b>Abbildung 5.44:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i12.	45
<b>Abbildung 5.45:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i10.	46
<b>Abbildung 5.46:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i23.	46
<b>Abbildung 5.47:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 6 nach dem Ausscheiden von fünf Items.	49
<b>Abbildung 5.48:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 6 nach dem Ausscheiden von fünf Items.	49
<b>Abbildung 5.49:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 6 nach dem Ausscheiden von fünf Items.	50
<b>Abbildung 5.50:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 vor dem Ausschluss von Aufgaben.	51
<b>Abbildung 5.51:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 9 vor dem Ausschluss von Aufgaben.	51
<b>Abbildung 5.52:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 9 vor dem Ausschluss von Aufgaben; die Ergebnisse werden nur unter Vorbehalt dargestellt.	52
<b>Abbildung 5.53:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 30.	52
<b>Abbildung 5.54:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 42.	53
<b>Abbildung 5.55:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 19.	53
<b>Abbildung 5.56:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 14.	54
<b>Abbildung 5.57:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 10 .	54

<b>Abbildung 5.58:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 9 nach dem Ausscheiden von fünf Items.	56
<b>Abbildung 5.59:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 9 nach dem Ausscheiden von fünf Items.	57
<b>Abbildung 5.60:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 9 nach dem Ausscheiden von fünf Items.	57
<b>Abbildung 5.61:</b> Grafische Modellkontrolle des Untertests 10 mittels des Teilungskriteriums Score im UT 2.	58
<b>Abbildung 5.62:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 vor dem Ausschluss von Aufgaben.	59
<b>Abbildung 5.63:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 11 vor dem Ausschluss von Aufgaben.	59
<b>Abbildung 5.64:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 11 vor dem Ausschluss von Aufgaben.	60
<b>Abbildung 5.65:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 29.	61
<b>Abbildung 5.66:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 6.	61
<b>Abbildung 5.67:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 11.	61
<b>Abbildung 5.68:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 7.	62
<b>Abbildung 5.69:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 46.	62
<b>Abbildung 5.70:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 11 nach dem Ausscheiden von fünf Items.	64
<b>Abbildung 5.71:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 11 nach dem Ausschluss von fünf Item.	64
<b>Abbildung 5.72:</b> Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 11 nach dem Ausscheiden von fünf Items.	65

## Tabellenverzeichnis

<b>Tabelle 2.1:</b> Items des AID 2 (Version 2.1 & Version 2.2) und der im AID 3 (Version 3.1) ausgeschlossenen Items; Darstellung auf Basis des Itempools der in der Vorgabe zur Normierung des AID3 (Version 1.1) Eingang fand.	7
<b>Tabelle 4.1:</b> Verteilung der Stichprobe auf das Alter in Jahren.	9
<b>Tabelle 4.2:</b> Die Verteilung der Stichprobe auf die unterschiedlichen Schultypen gereiht nach absteigenden Häufigkeiten.	10
<b>Tabelle 4.3:</b> Lagemaße und Streuungsmaß in T-Werten je Standarduntertest (UT1 bis UT11) des AID 2 (Version 2.1 und Version 2.2). Die Zusatzuntertests 5a, 5b und 10a werden in dieser Darstellung nicht berücksichtigt.	11
<b>Tabelle 5.1:</b> Darstellung der Ergebnisse der LQTs der Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 1 vor dem Ausschluss von Aufgaben.	15
<b>Tabelle 5.2:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 1 nach dem Ausschluss von sieben Aufgaben.	25
<b>Tabelle 5.3:</b> Überblick zu den Items des Untertests 1 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.	26
<b>Tabelle 5.4:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 2 vor dem Ausscheiden von Aufgaben.	28
<b>Tabelle 5.5:</b> Darstellung der Ergebnisse der LQTs für den UT 2 für die Teilungskriterien Score, Geschlecht und Muttersprache nach dem Ausschluss einer Aufgabe.	31
<b>Tabelle 5.6:</b> Überblick zu den Items des Untertests 2 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.	31
<b>Tabelle 5.7:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 3 vor dem Ausscheiden von Aufgaben.	33
<b>Tabelle 5.8:</b> Darstellung der Ergebnisse der LQTs für den UT 3 für die Teilungskriterien Score, Geschlecht und Muttersprache nach dem Ausschluss von zwei Aufgaben.	36
<b>Tabelle 5.9:</b> Überblick zu den Items des Untertests 3 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.	39
<b>Tabelle 5.10:</b> Darstellung des Ergebnisses des LQTs für das Teilungskriterium Score im Untertest 4 vor dem Ausschluss von Aufgaben.	39
<b>Tabelle 5.11:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 6 vor dem Ausschluss von Aufgaben.	41
<b>Tabelle 5.12:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 6 nach dem Ausschluss von sieben Aufgaben.	47

<b>Tabelle 5.13:</b> Überblick zu den Items des Untertests 6 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.	48
<b>Tabelle 5.14:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 9 vor dem Ausschluss von Aufgaben.	50
<b>Tabelle 5.15:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 9 nach dem Ausschluss von fünf Aufgaben.	55
<b>Tabelle 5.16:</b> Überblick zu den Items des Untertests 9 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.	55
<b>Tabelle 5.17:</b> Darstellung des Ergebnisses des LQTs des Untertests 10 für das Teilungskriterium Score im Untertest 2.	57
<b>Tabelle 5.18:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 11 vor dem Ausschluss von Aufgaben.	58
<b>Tabelle 5.19:</b> Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 11 nach dem Ausscheiden von fünf Aufgaben.	63
<b>Tabelle 5.20:</b> Überblick zu den Items des Untertests 11 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.	63

## **I. EINLEITUNG**

Zielsetzung der vorliegenden Diplomarbeit ist die Klärung der Fragestellung, ob der Ausschluss von einzelnen Aufgaben im Zuge einer vollständigen Überarbeitung und Neugestaltung eines psychologisch diagnostischen Verfahrens, basierend auf inhaltlichen Überlegungen a-posteriori auch testtheoretisch fundierbar ist. Den Ausgangspunkt stellen Items des Adaptiven Intelligenzdiagnostikums (AID) 2 (Version 2.1, Kubinger & Wurst, 2000; Version 2.2, Kubinger, 2009) dar, welche inhaltlich auffällig waren und aus diesem Grund im AID 3 (Version 3.1; Kubinger & Holocher-Ertl, 2014) ausgeschlossen wurden. Die Operationalisierung der Aufgabenstellung erfolgte mittels einer retrospektiven Modellprüfung unter Anwendung des dichotomen logistischen Modells von Rasch.

## **II. THEORETISCHER TEIL**

Sowohl die *Item-Response* Theorie (IRT) als auch das dichotome logistische Modell nach *Rasch* (1960, 1980), welche die Basis der vorliegenden Arbeit darstellen, wurden in der Literatur und im Rahmen von Diplomarbeiten am Arbeitsbereich für Psychologische Diagnostik der Psychologischen Fakultät der Universität Wien bereits mannigfach beschrieben. Aus diesem Grund kann an dieser Stelle von einer neuerlichen Beschreibung Abstand genommen werden kann. Für ein Studium der genannten methodischen Grundlagen sei beispielsweise auf Kubinger (1989a, 2003, 2006 und 2011), Fischer (1974) und Rasch (1980) verwiesen. Weiterführende Informationen zum Adaptiven Intelligenzdiagnostikum finden sich bei Kubinger und Wurst (2000), Kubinger (2009) sowie Kubinger und Holocher-Ertl (2014).

### **1. Vorteile eines Tests, der dem dichotomen logistischen Modell von *Rasch* entspricht**

Wenn die Items eines psychologisch diagnostischen Verfahrens dem dichotomen logistischen Modell nach *Rasch* (oder einem anderen Modell der IRT) entsprechen zählt neben der Verrechnungsfairness auch die Anwendbarkeit des adaptiven Testens zu den resultierenden Vorteilen. Die Modellgültigkeit erlaubt nämlich auch Tpn zueinander in Relation zu setzen, die unterschiedliche Items bearbeitet haben.

Wenn adaptiv getestet wird, werden jene Items vorgegeben, welche der Fähigkeit der Tp am besten entsprechen. Der Informationsgewinn ist dann maximal, wenn Itemschwierigkeit und Personenfähigkeit gleich groß sind. Je mehr Informationen über einen Parameter bekannt sind, desto genauer ist seine Schätzung. Das adaptive Testen stellt damit ein optimales Verhältnis zwischen Testlänge und Messgenauigkeit her (Kubinger, 2003).

*Branched-testing* ist jene Art des adaptiven Testens, welche beim AID zu Anwendung gelangt. Dabei werden Items in Gruppen zusammengefasst. Die Anzahl der gelösten Items einer Testperson (Tp) innerhalb einer Itemgruppe bestimmt welche Itemgruppe danach vorgegeben wird. Zuvor bestimmte Regeln legen fest, ob nachfolgende Items schwieriger, leichter oder nahezu gleichschwierig sind (Kubinger, 1989b).

## 2. Das Adaptive Intelligenzdiagnostikum (AID)

Wie in der Einleitung erwähnt, beschränkt sich die Beschreibung des AID im Rahmen dieser Diplomarbeit auf das Wesentliche. Es wird kein Anspruch auf Vollständigkeit erhoben. Eine detaillierte Darstellung des AID 2 (Version 2.1) findet sich bei Kubinger und Wurst (2000), jene des AID 2 (Version 2.2) bei Kubinger (2009).

Das Adaptive Intelligenzdiagnostikum basiert auf dem Testkonzept von Wechsler. Der AID ist eine Intelligenztestbatterie für Kinder und Jugendliche im Alter von 6;0 bis 15;11 Jahren. Sie findet Einsatz in der Schulpsychologie, der Berufs- und Bildungsberatung und in der klinischen Psychologie (Kubinger & Wurst, 2000).

Dem AID „liegt der pragmatische Standpunkt zugrunde, ziemlich viele (komplexe und basale) Fähigkeiten, die für 'intelligentes' Verhalten verantwortlich scheinen, zu erfassen“ (Kubinger & Wurst, 2000, S. 14). Von diesem Ansatz ausgehend kann „'Intelligenz' definiert werden als: das Bündel aller kognitiven Voraussetzungen, die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln“ (Kubinger & Wurst, 2000, S. 30).

Aufbauend auf dieser Intelligenztheoretischen Sichtweise und untermauert durch teststatistische Befunde (vgl. Kubinger & Wurst, 2000) wird die Generalfaktorentheorie der Intelligenz sensu *Spearman*, endgültig verworfen. Dafür spricht die empirisch gefundene Struktur des AID, welche sich in einer vier-Faktorenlösung beschreiben lässt: "Informationsverarbeitung in der gesellschaftlichen Umwelt", "Informationsverarbeitung neuer Inhalte", "Auffassungskapazität" und "(Re-)Produktionsfähigkeit durch Strukturierung" (Kubinger & Wurst, 2000, S. 14).

Es wird auch notwendig, das Kompensationsmodell, welches dem Intelligenzquotienten (IQ) zugrunde liegt, zu verwerfen. Fähigkeitsdefizite in bestimmten Bereichen können nicht durch ein 'Mehr' an Fähigkeiten in anderen Bereichen ausgeglichen werden.

Im AID wird ein Defizitmodell der Intelligenz propagiert. Kubinger und Wurst (2000) beziehen sich dabei auf das Gleichnis, dass „jede Kette nur so stark ist wie ihr schwächstes Glied“ (S. 31) sei. Für den AID konnte „eine Polarisierung in förderungsabhängige und (eher) förderungsunabhängige Untertests empirisch nachgewiesen werden“ (Kubinger & Wurst, 2000, S. 14). Mittels der durch den AID aufgedeckten Fähigkeitsdefizite wird somit die Möglichkeit zur spezifischen Förderung gegeben.

Bezüglich der inhaltlichen Gestaltung der Untertests des AID sei hervorgehoben, dass ihnen stets operationale Definitionen zugrunde liegen. „Es wurde begrifflich exakt festgelegt, was mit ihnen jeweils zu erfassen beabsichtigt war“ (Kubinger & Wurst, 2000, S. 15). Ausgehend von den operationalen Definitionen können die entwickelten Untertests zwei Fähigkeitsbereichen zugeordnet werden. Einerseits erfassen sie verbal-akustische Fähigkeiten. Darunter werden Aufgabenstellungen verstanden, „die sich auf akustisches Auffassen und verbalisiertes Agieren beziehen“ (Kubinger, 2009, S. 6). Andererseits werden manuell-visuelle Fähigkeiten gemessen. Letztere Aufgaben erfordern die Handhabung von Gegenständen sowie die Erfassung visueller Gestalten (Kubinger & Wurst, 2000).

## 2.1 Aufbau des AID

Die beiden Versionen des AID 2 (Version 2.1, Kubinger & Wurst, 2000; Version 2.2, Kubinger, 2009) bestehen jeweils aus insgesamt 11 Untertests und zusätzlich drei Zusatztests. Die beiden Versionen unterscheiden sich hinsichtlich der Normierungsstichproben, nicht jedoch bezüglich der Items. Acht der elf Standarduntertests werden adaptiv im Sinne des *branched-testing* vorgegeben.

Zu diesen acht Subtests zählen die Untertests 1 (Alltagswissen), 3 (Angewandtes Rechnen), 4 (Soziale und Sachliche Folgerichtigkeit), 6 (Synonyme Finden), 8 (Antizipieren und Kombinieren - figural), 9 (Funktionen Abstrahieren), 10 (Analysieren und Synthetisieren - abstrakt) und 11 (Soziales Erfassen und Sachliches Reflektieren).

Im Rahmen dieser Arbeit lag das Hauptaugenmerk auf jenen Aufgaben des AID 2 (Version 2.1 und Version 2.2) welche aus inhaltlichen Gründen nicht in den AID 3 (Version 3.1) übernommen wurden. Tabelle 2.1 gibt einen Überblick über diese Items. Es wird an dieser Stelle darauf hingewiesen, dass diese Diplomarbeit eine retrospektive testtheoretische Arbeit darstellt. Die Ergebnisse fließen nicht in die aktuelle, Publikation des AID 3 (Version 3.1.) ein. Außerdem ist darauf hinzuweisen, dass auf Basis neuester wissenschaftlicher Erkenntnisse (Kubinger, K. D., Steinfeld, J. & Reif M., in Vorb.) alle *Likelihood-Quotienten-Tests* nach *Andersen* (LQTs) nicht signifikant sind und daher gar kein Item auszuschließen ist. Diese Erkenntnis betrifft alle in dieser Diplomarbeit vorgenommenen Analysen mittels des LQTs für jeden Untertest (vgl. Kapitel 5).

**Tabelle 2.1:** Items des AID 2 (Version 2.1 & Version 2.2) und der im AID 3 (Version 3.1) ausgeschlossenen Items.

		<b>Im AID 3</b>	<b>Anzahl</b>
	<b>Items gesamt</b>	<b>ausgeschlossene</b>	<b>ausgeschlossener</b>
		<b>Items</b>	<b>Items</b>
UT 1	60	i1, i4, i8, i11, i22, i24, i28, i29, i32, i34, i35, i36, i37, i39, i40, i42, i47, i49, i50, i54, i57, i58, i59, i60	24
UT 2	12	i8, i9, i12	3
UT 3	60	i2, i6, i24, i26, i36, i37, i39, i40, i41, i48, i50, i53	12
UT 4	18		0
UT 5	48		0
UT 6	60	i4, i5, i9, i10, i11, i12, i13, i14, i20, i21, i22, i24, i25, i26, i29, i30, i31, i32, i33, i34, i35, i36, i37, i40, i41, i46, i47, i48, i49, i50, i51, i53, i57, i59	34
UT 7	105		0
UT 8	11		0
UT 9	60	i6, i17, i18, i22, i29, i31, i33, i35, i39, i45, i50, i53, i57	13
UT 10	22	I0, i00, i13, i14	4
UT 11	60	i3, i4, i12, i18, i19, i21, i22, i24, i28, i29, i30, i35, i39, i42, i43, i46, i53, i55, i58, i59	20

### III. EMPIRISCHER TEIL

#### 3. Ziel der Untersuchung

Ziel dieser Diplomarbeit war, retrospektiv zu zeigen, dass der Ausschluss von Aufgaben, welche im AID 2 (Version 2.1, Kubinger & Wurst, 2000 und Version 2.2, Kubinger, 2009) noch zur Anwendung gelangten, auf Basis inhaltlicher Überlegungen für den AID 3 (Version 3.1, Kubinger & Holoher-Ertl, 2014), auch teststatistisch sinnvoll ist. Diese Überprüfung wurde unter den Prämissen der *Item-Response-Theorie* mittels der Anwendung des dichotomen logistischen Modells nach *Rasch* nach dem *state of the art* (Kubinger, 2005) teststatistisch durchgeführt, "um den Grad der Bewährung des *Rasch*-Modells als ausreichend qualifizieren zu können" (Kubinger, 2005).

Einschränkend muss auf Basis neuester wissenschaftlicher Erkenntnisse (Kubinger, K. D., Steinfeld, J. & Reif M., in Vorb.) darauf hingewiesen werden, dass alle *Likelihood*-Quotienten-Tests nach *Andersen* nicht signifikant sind und daher gar kein Item auszuschließen ist. Diese Erkenntnis betrifft alle in dieser Diplomarbeit vorgenommenen Analysen mittels des LQTs für jeden Untertest.

#### 4. Stichprobe

Die Stichprobe setzt sich aus Tpn zusammen, welche aus Eigeninitiative oder über eine Zuweisung einer psychologischen Diagnostik durch zwei erfahrene klinische und Gesundheitspsychologinnen getestet wurden. Die psychologisch diagnostischen Testungen wurden mittels des AID 2 (Version 2.1, Kubinger & Wurst, 2000; Version 2.2, Kubinger, 2009) durchgeführt. Die deskriptive Beschreibung der vorliegenden Stichprobe umfasst die Variablen Geschlecht, Alter, Muttersprache, Schulform und durchschnittlich erzielte Leistungen im AID.

Die Testungen wurden im Zeitraum von März 2001 bis Dezember 2013 durchgeführt. Die Stichprobe ( $n=445$ ) setzt sich aus 63,1% Buben und 36,9% Mädchen zusammen. Zwei Kinder (0,4%) stammen aus dem Burgenland, 131 (29,4%) aus Niederösterreich und 312 (70,1%) aus Wien. Zur deskriptiven Beschreibung wurden die Daten mit SPSS Version 20 ausgewertet.

#### 4.1.1 Alter

Die Testpersonen der Stichprobe weisen ein Alter zwischen 6 und 15 Jahren ( $Md_{\text{Alter}} = 9$  Jahre) auf. Tabelle 4.1 gibt einen Überblick bezüglich der Verteilung der Altersgruppen im vorliegenden Datensatz.

**Tabelle 4.1:** Verteilung der Stichprobe auf das Alter in Jahren ( $n=445$ ).

<b>Alter in Jahren</b>	<b>Absolute Häufigkeit</b>	<b>%</b>
6	60	13,5
7	87	19,6
8	61	13,7
9	48	10,8
10	55	12,4
11	38	8,5
12	32	7,2
13	19	4,3
14	27	6,1
15	18	4,0

---

$Md_{\text{Alter}} = 9$  Jahre

### 4.1.2 Muttersprache

Bei 357 (80,2%) der Kinder ist die Muttersprache Deutsch, bei 88 (19,8%) eine andere Sprache als Deutsch. Bei Kindern mit einer anderen Muttersprache als Deutsch waren am häufigsten Serbisch, Bosnisch oder Kroatisch mit 34 Kindern (7,6%) vertreten. Die zweithäufigsten nicht deutschen Muttersprachen stellen Türkisch und Arabisch mit jeweils 9 Kindern (2,0%) dar.

### 4.1.3 Schulform

Mehr als die Hälfte der Tpn besuchte zum Testzeitpunkt eine Volksschule (62,5%). Die zweitgrößte Gruppe stellen mit 121 Tpn (27,2%) Kinder dar, welche eine Allgemeinbildende Höhere Schule absolvieren. Eine detaillierte Darstellung der Verteilung der Tpn auf die Schultypen kann Tabelle 4.2 entnommen werden.

**Tabelle 4.2:** Die Verteilung der Stichprobe auf die unterschiedlichen Schultypen gereiht nach absteigenden Häufigkeiten ( $n=445$ ).

<b>Schultyp</b>	<b>Absolute Häufigkeit</b>	<b>%</b>
Volksschule	278	62,5
Allgemeinbildende höhere Schule	121	27,2
Kooperative Mittelschule	23	5,2
Hauptschule	17	3,8
Polytechnikum	3	0,7
Berufsbildende höhere Schule	3	0,7

#### 4.1.4 Durchschnittlich erzielte Leistungen

Tabelle 4.3 gibt die durchschnittlichen, altersnormierten T-Werte je Untertest sowie deren Standardabweichung und die minimalen beziehungsweise maximalen T-Werte wider.

**Tabelle 4.3:** Lagemaße und Streuungsmaß in T-Werten je Standarduntertest (UT1 bis UT11) des AID 2 (Version 2.1 und Version 2.2). Die Zusatzuntertests 5a, 5b und 10a werden in dieser Darstellung nicht berücksichtigt ( $n=445$ ).

	$\bar{y}$	S	Minimum	Maximum	n
UT 1	50,35	9,57	19	81	443
UT 2	47,13	9,16	19	73	444
UT 3	51,12	12,18	17	81	443
UT 4	54,49	11,54	19	81	444
UT 5v	53,28	8,09	27	81	443
UT 5r	49,85	9,21	19	82	443
UT 6	51,68	9,33	19	81	442
UT 7km	51,58	9,17	20	81	443
UT 7ass	48,93	8,96	18	79	443
UT 8	48,93	10,31	23	81	443
UT 9	50,74	9,57	21	77	441
UT 10	50,12	11,16	19	81	442
UT 11	50,55	11,24	19	81	440

## 5. Ergebnisse

Die *Rasch*-Modell-Reanalysen des AID 2 (Version 2.1 und Version 2.2) wurden mit dem Programmpaket eRm (Version 0.15-4; Maier, Hatzinger & Mair, 2014) für die *Open-Source* Statistik-Software R durchgeführt. Von der Analyse ausgenommen wurden die Zusatztests (5a, 5b und 10a) sowie die Untertests 5, 7 und 8. Dabei wird analog zur Analyse des AID 2 (Version 2.2, Kubinger, 2009) vorgegangen: „Die Untertests 5 [...] und 7 [...] müssen keiner testtheoretischen Analyse unterzogen werden, um die Verrechnungen der Testleistungen zu Testwerten als fair zu statuieren“ (S.48). Für den Untertest 8 sieht das Verrechnungsschema eine dreikategorielle Verrechnung vor die nach Kubinger (2009) mit dem „*Partial-Credit-Modell* von *Masters*“ zu analysieren ist. Im Rahmen dieser Diplomarbeit wurde die Reanalyse aber auf jene Untertests eingeschränkt, deren Analyse mit dem dichotomen logistischen Modell nach *Rasch* möglich ist, wodurch der Untertest 8 nicht einfließen konnte.

Bei den Untersuchungen wurde im Sinne des *state of the art* nach Kubinger (2005) vorgegangen: im Fall eines signifikanten Ergebnisses von *Andersens Likelihood* Quotienten-Tests sollen unter Zuhilfenahme der Grafischen Modellkontrolle nach Rasch (vgl. beispielsweise Kubinger, 1989a) jene Aufgaben identifiziert werden, welche für die Modellabweichung verantwortlich sind. Ziel der Analyse ist solange schrittweise Aufgaben auszuschließen, bis zumindest a-posteriori *Rasch*-Modell Konformität erreicht wurde. Bei den Grafischen Modellkontrollen wurde, geschuldet der Übersichtlichkeit, in Anlehnung an das Manual des AID 3 (Version 1.1, Kubinger & Holocher-Ertl, 2014) von einer Beschriftung der Aufgaben abgesehen.

Eine Abweichung ist nach Kubinger (2005) dann als praktisch relevant anzusehen, wenn die Differenz der Schätzungen der Parameter aus zwei Teilstichproben mehr als ein Zehntel der Spannweite aller Itemparameterschätzungen beträgt.

Folgende Teilungskriterien wurden für die Auswertung herangezogen:

Internes Teilungskriterium:

- Anzahl gelöster Aufgaben (Rohwert kleiner gleich Median vs. Rohwert größer Median)

Externe Teilungskriterien:

- Geschlecht
- Muttersprache (Deutsch als Muttersprache - nichtdeutsche Muttersprache)
- Alter (6- bis 9-Jährige vs. 10- bis 15-Jährige)

Einschränkend ist anzumerken, dass aufgrund der Art der Realisierung des *branched-testings* die Untertests 1, 3, 6, 9 und 11 nicht anhand des Teilungskriteriums Alter analysiert werden konnten. Die vom Alter abhängige Auswahl des ersten Aufgabenblocks führt dazu, dass in den Teilstichproben bei einer Teilung nach dem Alter zu wenige Aufgaben in beiden Gruppen gemeinsam vorkommen (Kubinger, 2009).

Zur Analyse der Untertests 4 und 10 ist anzumerken, dass es wie im Manual des AID 2 (Version 2.1, Kubinger & Wurst, 2000) diskutiert, aufgrund der gewählten Verzweigungsschemata in beiden Untertests zu methodischen Artefakten kommt. Aus diesem Grund wird analog zum Vorgehen von Kubinger und Holocher-Ertl (AID 3, Version 3.1; 2014) für die Analyse der beiden Untertests als Teilungskriterium der Median des Scores im Untertest 2 herangezogen.

Hinsichtlich des Vorgehens bei den Analysen ist Folgendes anzumerken: es konnten nicht immer alle vorgegebenen Aufgaben und Testpersonen mittels des dichotomen logistischen Modells nach *Rasch* untersucht werden, da manche Items entweder von allen Testpersonen gelöst oder nie gelöst wurden. Auch jene Aufgaben konnten nicht analysiert werden, welche durch Teilung der Stichprobe anhand der gewählten Teilungskriterien zumindest in einer der betreffenden Teilstichproben immer oder nie gelöst wurden. Weiters zeigten manche Aufgaben zu extreme Lösungshäufigkeiten (Lösungshäufigkeiten über 95% beziehungsweise unter 5%) und wurden nicht analysiert. Die Anzahl schätzbarer Items ist je Untertest anhand der Anzahl der angegebenen Freiheitsgrade ( $df+1$ ) in den dargestellten Tabellen ersichtlich.

Das Risiko 1. Art wurde mit  $\alpha=0,05$  festgesetzt. Um jedoch der Kumulierung des Risikos 1. Art bei Anwendung des LQT für die vier Teilungskriterien entgegenzuwirken wurde das Signifikanzniveau für die Annahme der drei Teilungskriterien sensu Kubinger (2005) auf  $\alpha=0,01$  adjustiert.

Es muss darauf hingewiesen werden, dass jene Items welche im AID 3 (Version 3.1) zur Anwendung gelangen aus Gründen der Testsicherheit inhaltlich nicht widergegeben werden können.

Abschließend muss auf aktuelle Entwicklungen eingegangen werden. Die jüngsten Ergebnisse von Kubinger et al. (in Vorb.) zeigen, nämlich, dass alle *Likelihood*-Quotienten-Tests nach *Andersen* (LQTs) nicht signifikant sind und daher darf gar kein Item ausgeschieden werden. Diese neuen Erkenntnisse sind unbedingt bei der Betrachtung der dargestellten Ergebnisse zu beachten.

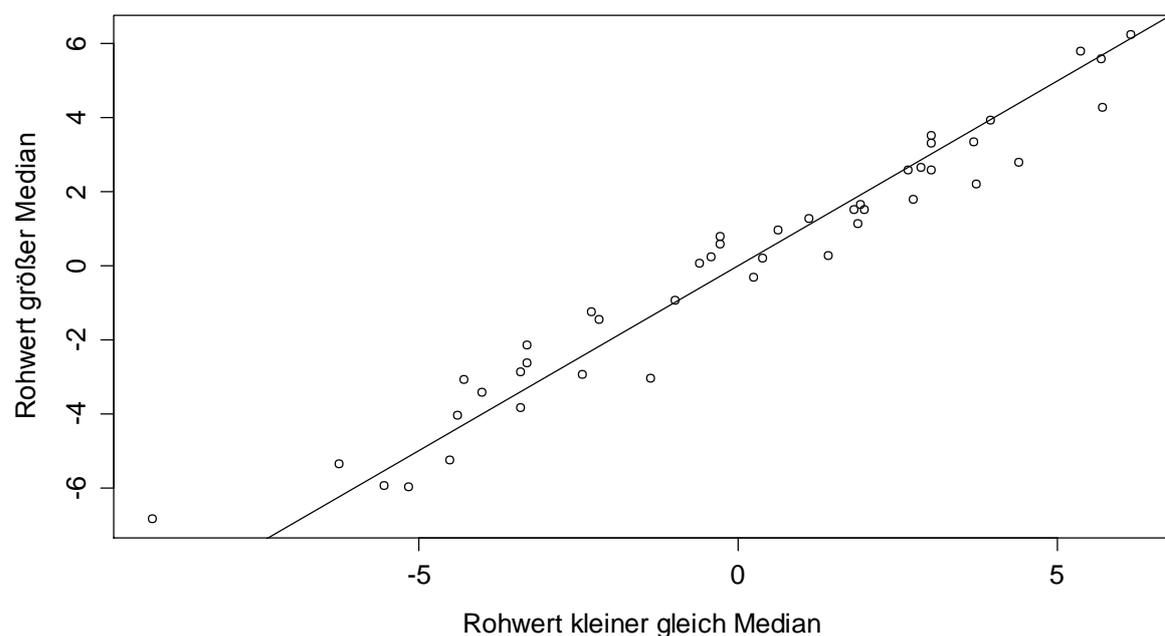
## 5.1 Untertest 1 Alltagswissen

Im UT 1 mussten zwei Testpersonen (0221 und 0017) von der Analyse ausgeschlossen werden. Folgende Items konnten in der erhobenen Stichprobe nicht analysiert werden: i1, i2, i3, i4, i5, i6, i7, i8, i9, i10, i14, i27, i31, i32, i33, i34, i35, i36, i37, i38, i39, i40, i46 und i56d. Tabelle 5.1 kann entnommen werden, dass es für alle drei Teilungskriterien zu signifikanten Modellabweichungen gekommen ist. Es ist anzumerken, dass insgesamt elf der vierundzwanzig nicht analysierbaren Aufgaben *Linking-Items* sind.

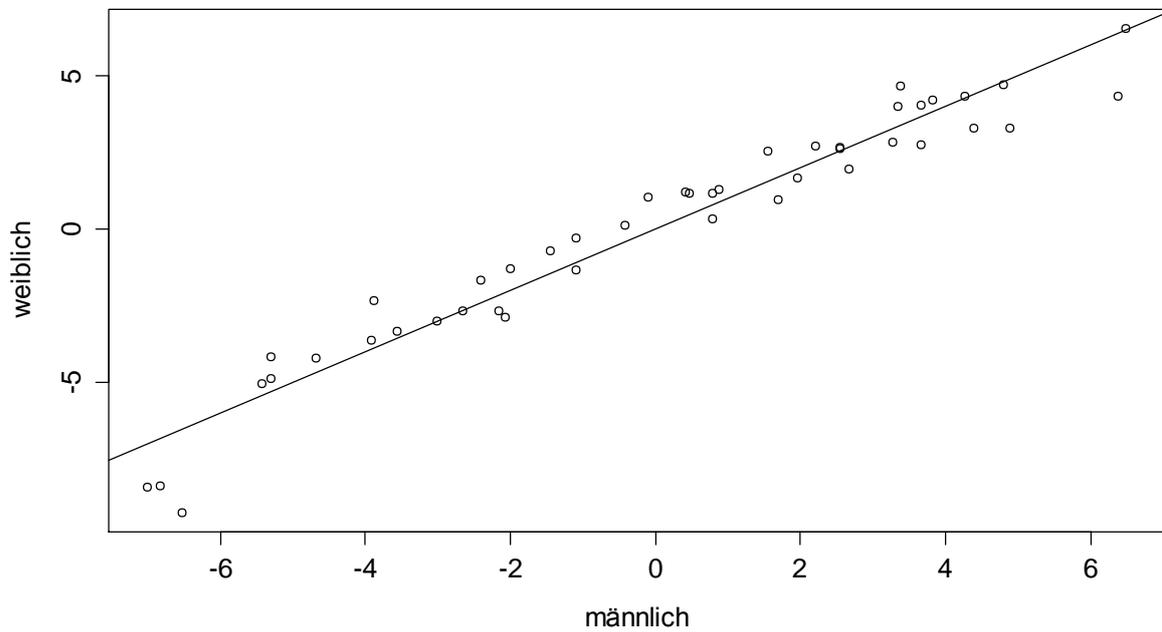
**Tabelle 5.1:** Darstellung der Ergebnisse der LQTs der Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 1 vor dem Ausschluss von Aufgaben ( $n=443$ ).

Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	p
Score	80,300	43	67,459	<0,001
Geschlecht	89,827	44	68,710	<0,001
Muttersprache	124,491	41	64,950	<0,001

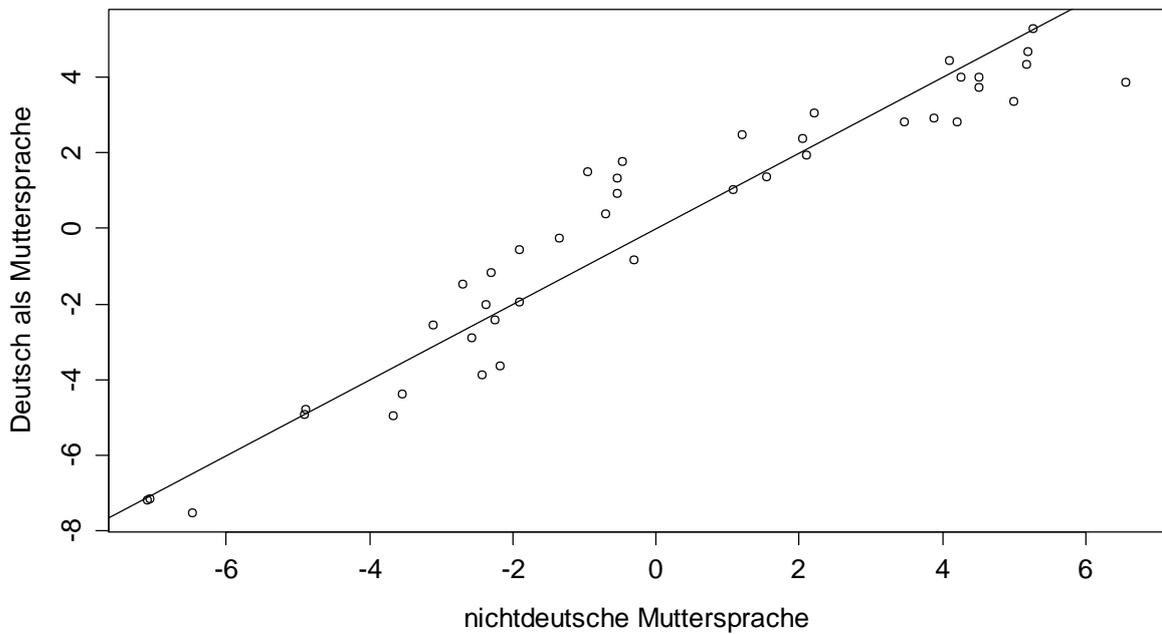
Die Abbildungen 5.1 bis 5.3 zeigen die Grafischen Modellkontrollen der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 1 vor dem Ausschluss von Items.



**Abbildung 5.1:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 vor dem Ausschluss von Aufgaben (n=443).

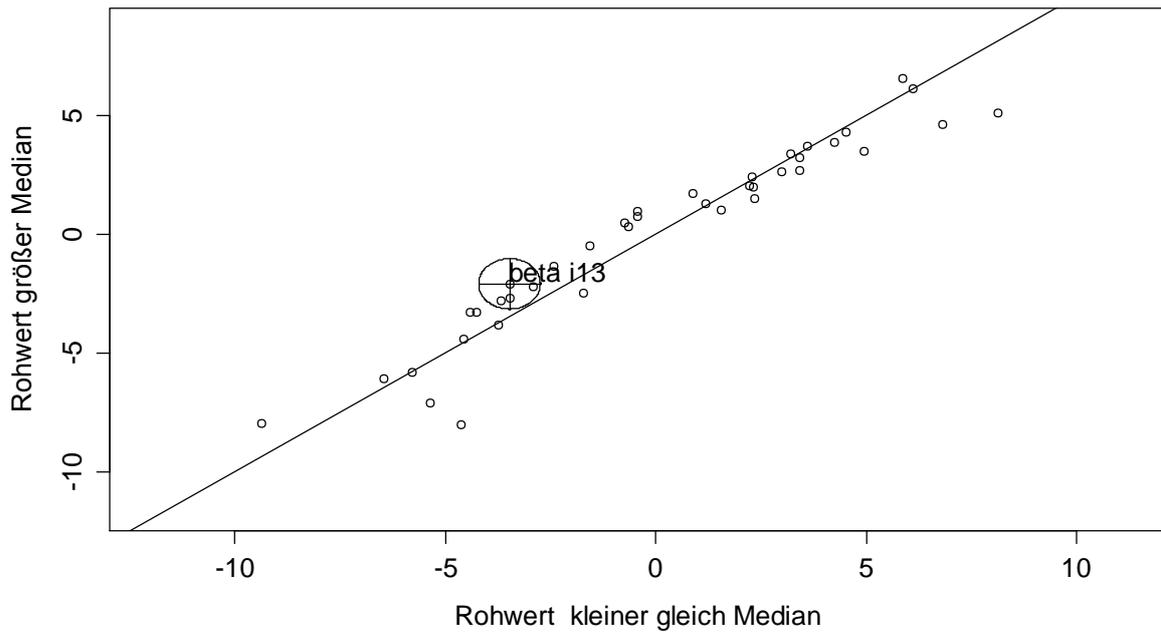


**Abbildung 5.2:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 1 vor dem Ausschluss von Aufgaben (n=443).

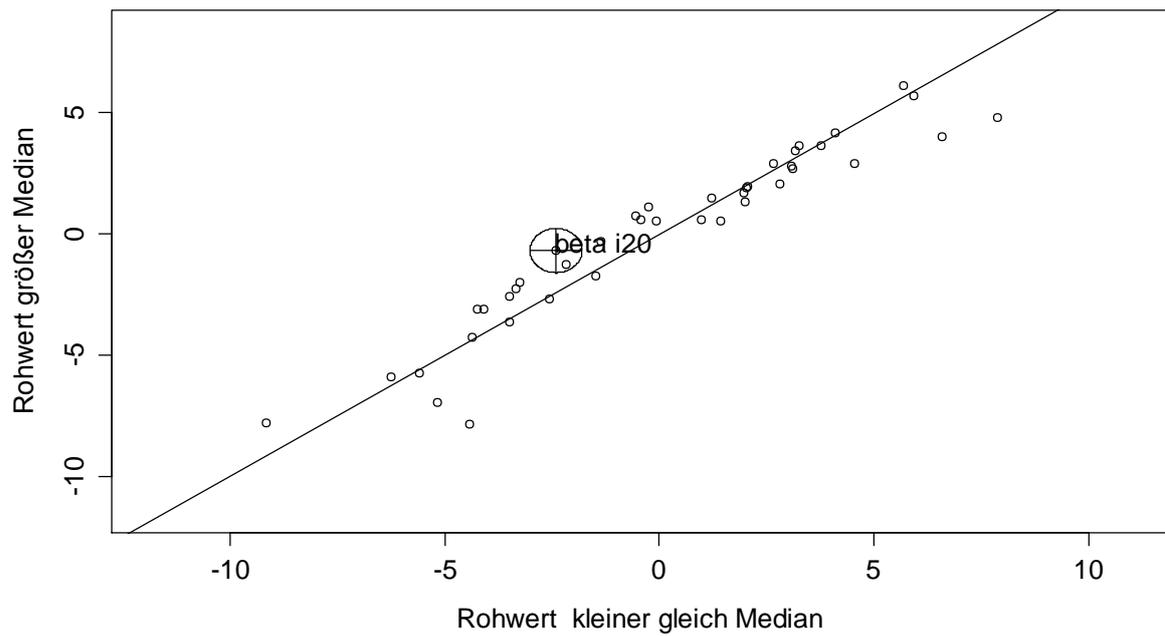


**Abbildung 5.3:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 vor dem Ausschluss von Aufgaben; (n=443).

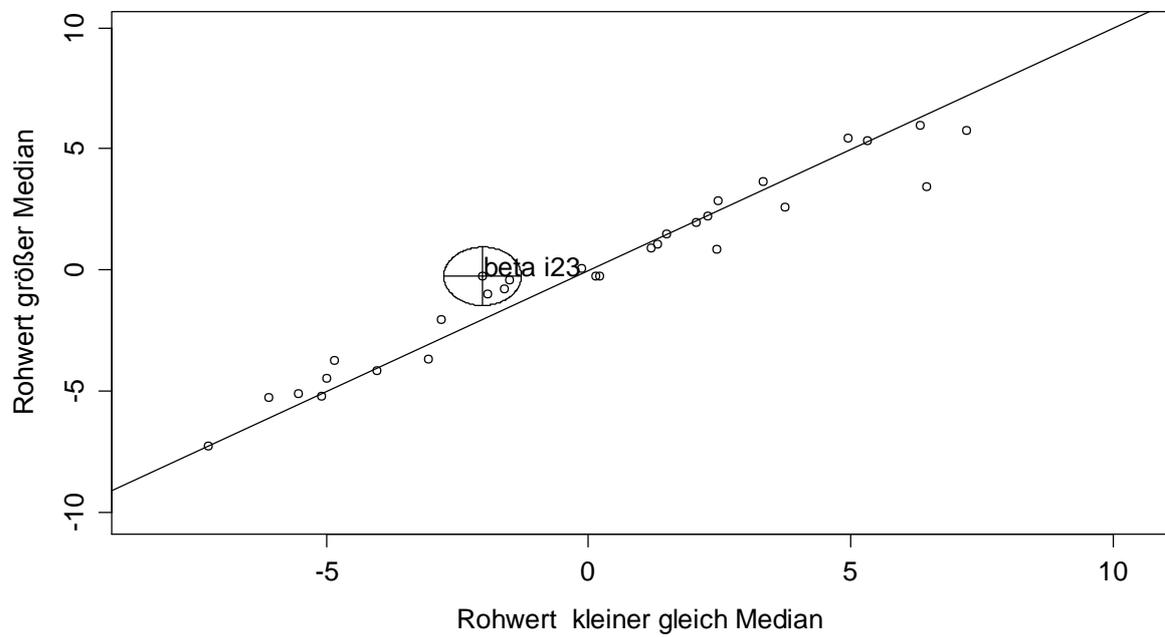
Die Grafischen Modellkontrollen nach *Rasch* für das Teilungskriterium Score zeigten, dass die Items i13, i20, i23, i31, i42 und i57 auffällig waren. Sie wurden aufgrund der Differenz der Parameterschätzungen in den LQTs von -1,382, -1,734, 1,766, -3,054, 4,191 und -1,792 ausgeschlossen (vgl. Abbildung 5.4 bis 5.9).



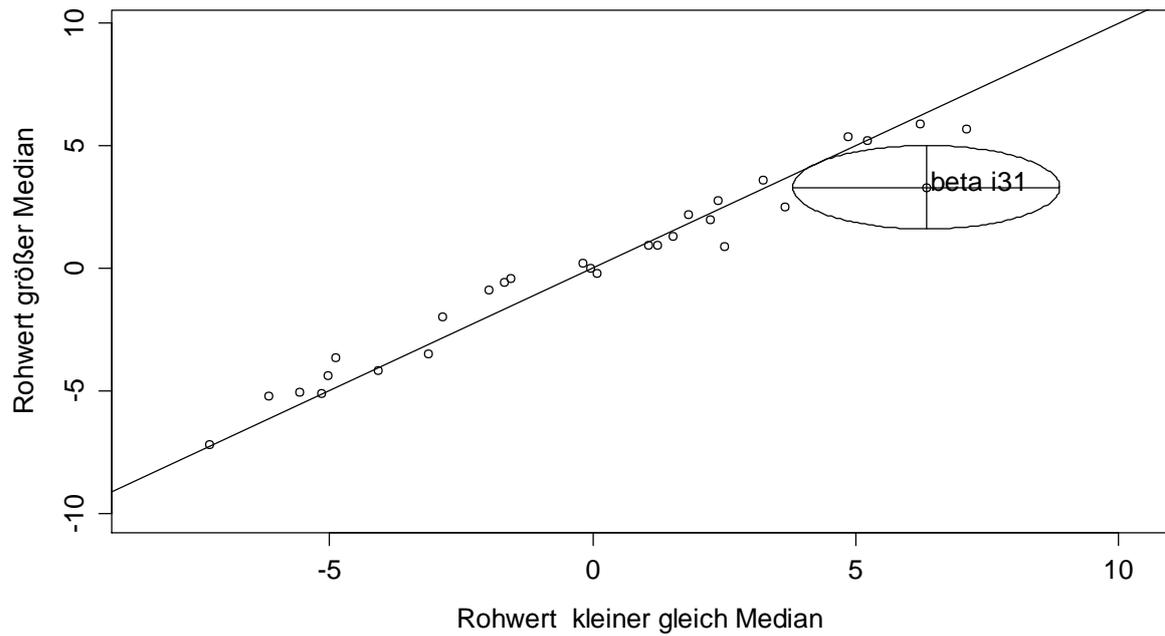
**Abbildung 5.4:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i13 ( $\alpha=0,01$ ); ( $n=443$ ).



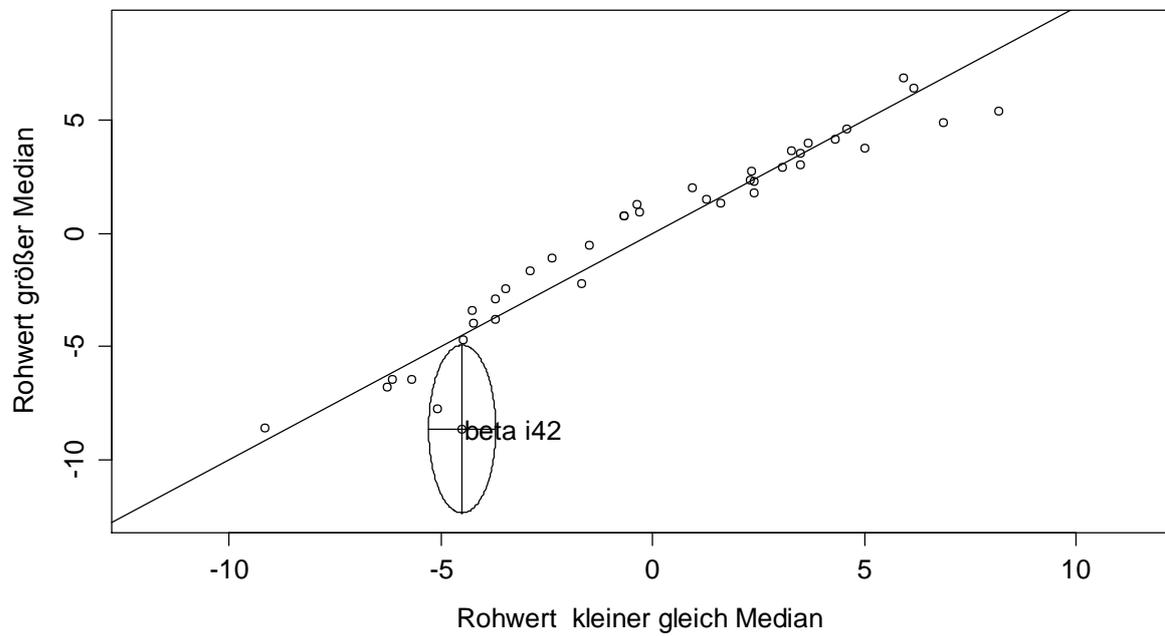
**Abbildung 5.5:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i20 ( $\alpha=0,01$ ); ( $n=443$ ).



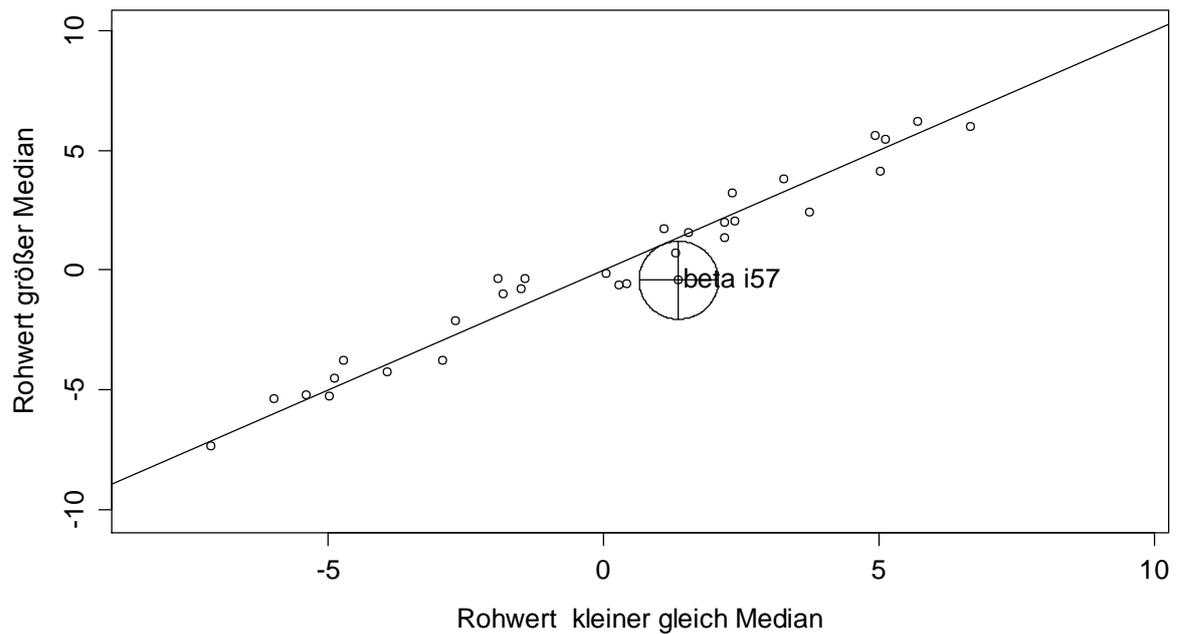
**Abbildung 5.6:** Grafische Modellkontrolle des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i23 ( $\alpha=0,01$ ); ( $n=443$ ).



**Abbildung 5.7:** Grafische Modellkontrolle des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i31 ( $\alpha=0,01$ ); ( $n=443$ ).

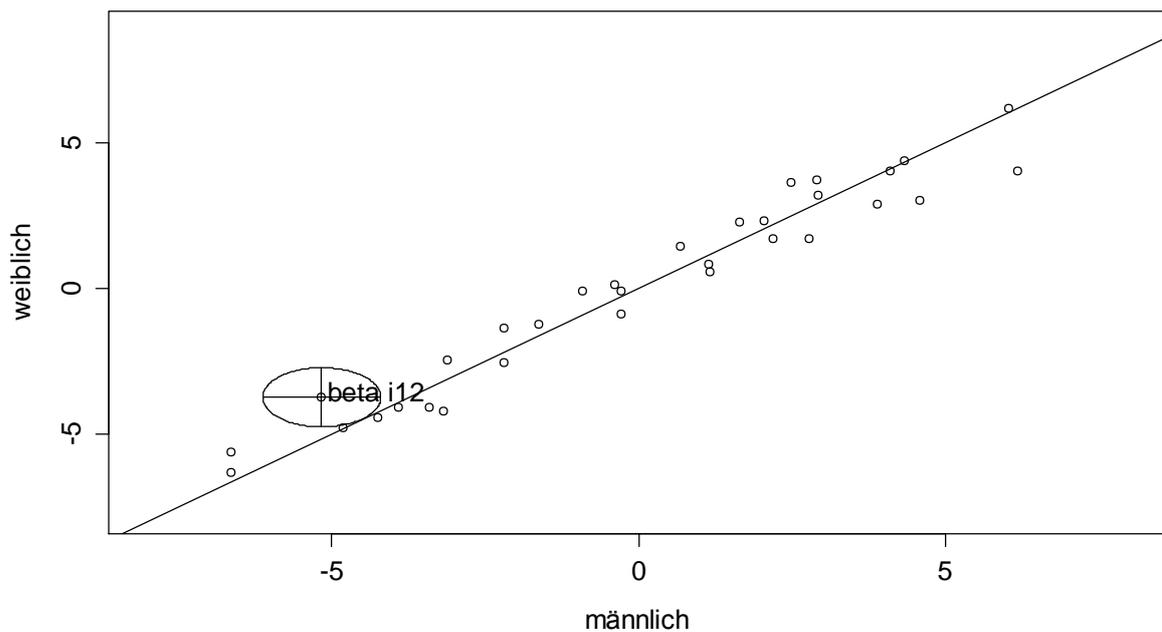


**Abbildung 5.8:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i42 ( $\alpha=0,01$ ); ( $n=443$ ).

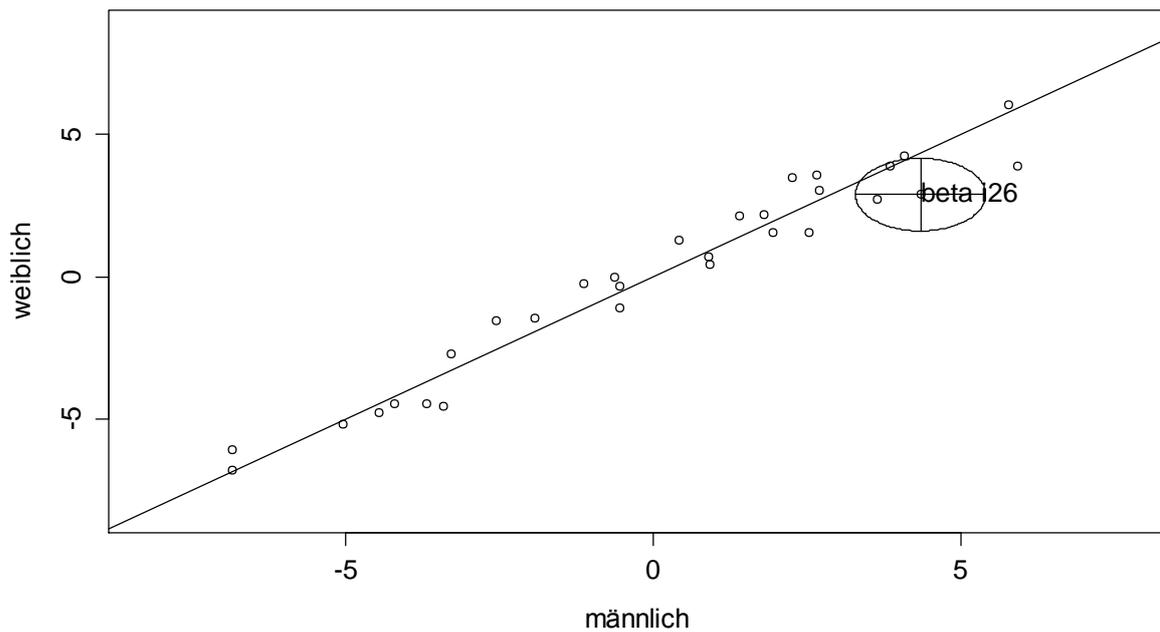


**Abbildung 5.9:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 1 mit Standardschätzfehler des Items i57 ( $\alpha=0,01$ ); ( $n=443$ ).

Bei der Grafischen Modellkontrolle für das Teilungskriterium Geschlecht waren die Aufgaben i12 und i26 auffällig und wurden ausgeschlossen (vgl. Abbildungen 5.10 und 5.11).

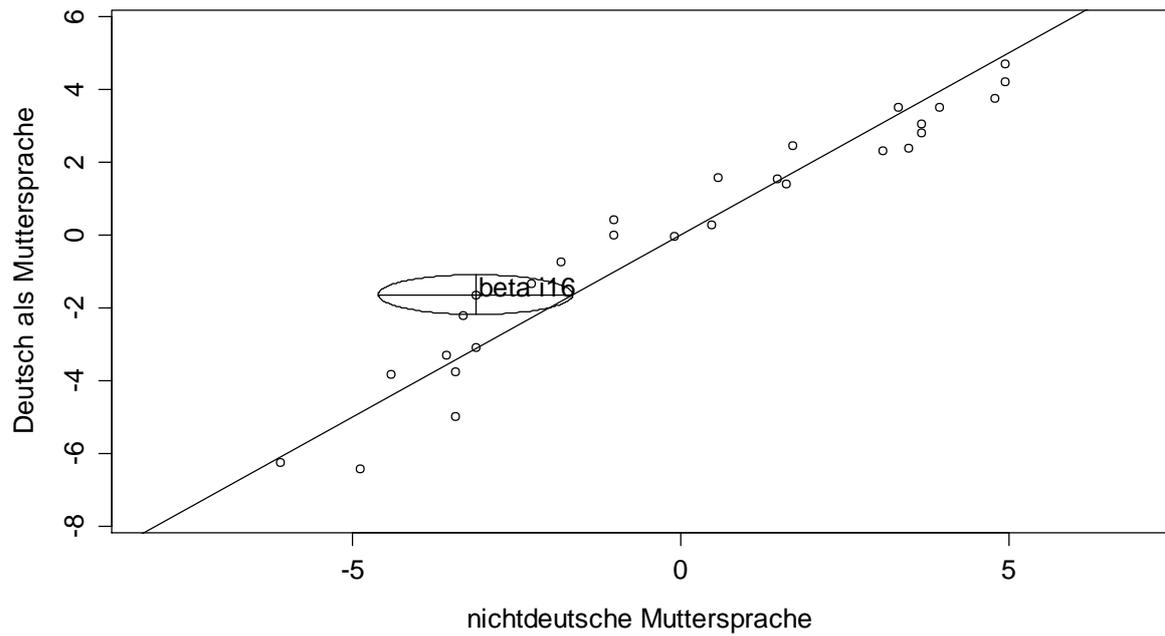


**Abbildung 5.10:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 1 mit Standardschätzfehler des Items i12 ( $\alpha=0,01$ ); ( $n=443$ ).

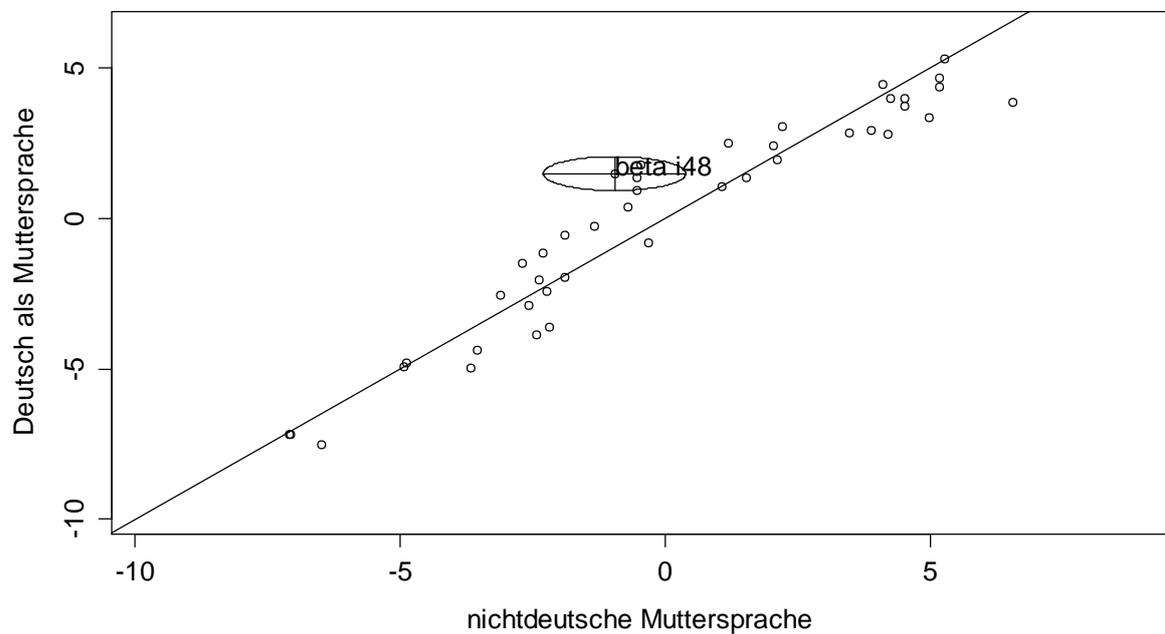


**Abbildung 5.11:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 1 mit Standardschätzfehler des Items i26 ( $\alpha=0,01$ ); ( $n=443$ ).

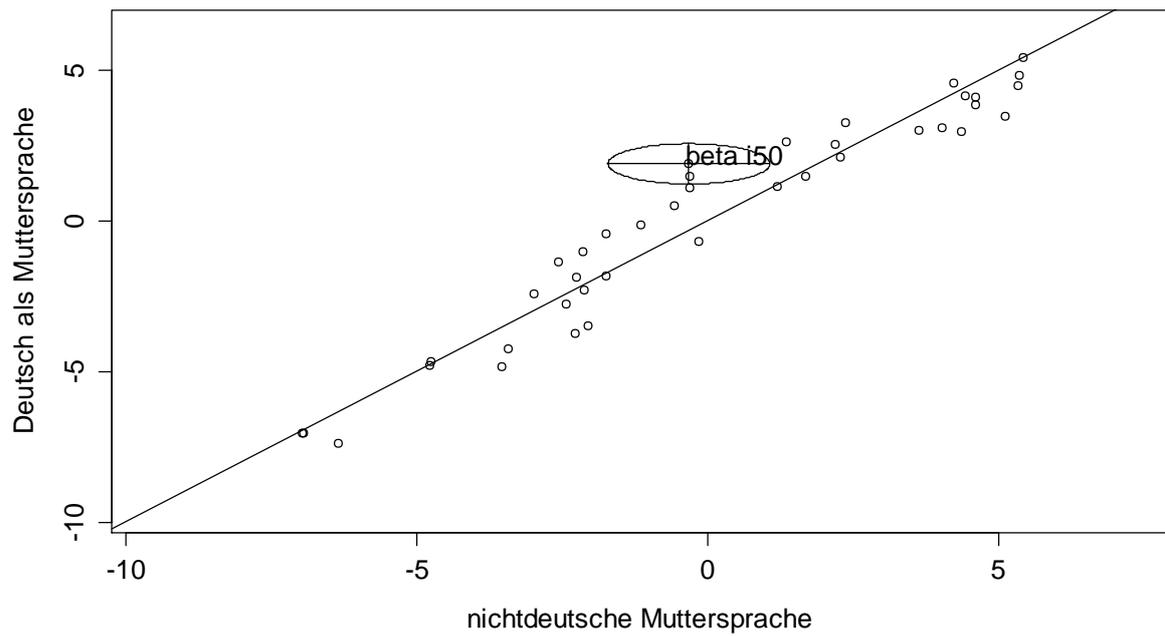
Bei den Grafischen Modellkontrollen für das Teilungskriterium Muttersprache waren die Items i16, i48, i50, i51, i54 und i55 auffällig und wurden aufgrund der Differenz der Parameterschätzungen in den LQTs von 1,472, 2,283, 2,223, 2,693, 1,852 und 1,584 ausgeschlossen (vgl. Abbildung 5.12 bis 5.17).



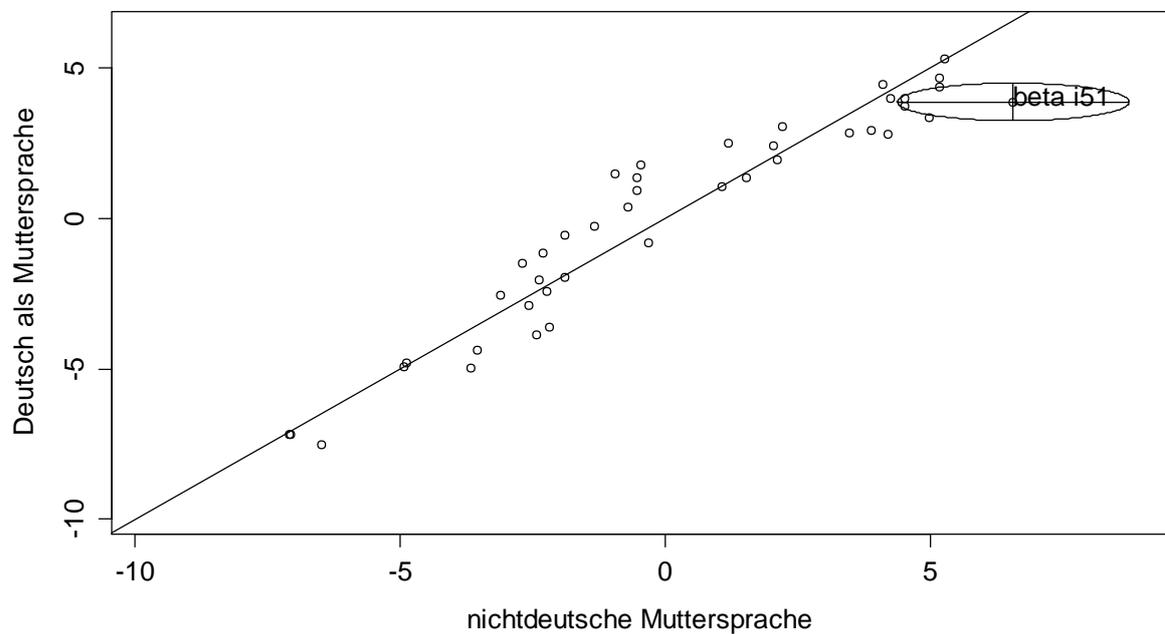
**Abbildung 5.12:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i16 ( $\alpha=0,01$ ); ( $n=443$ ).



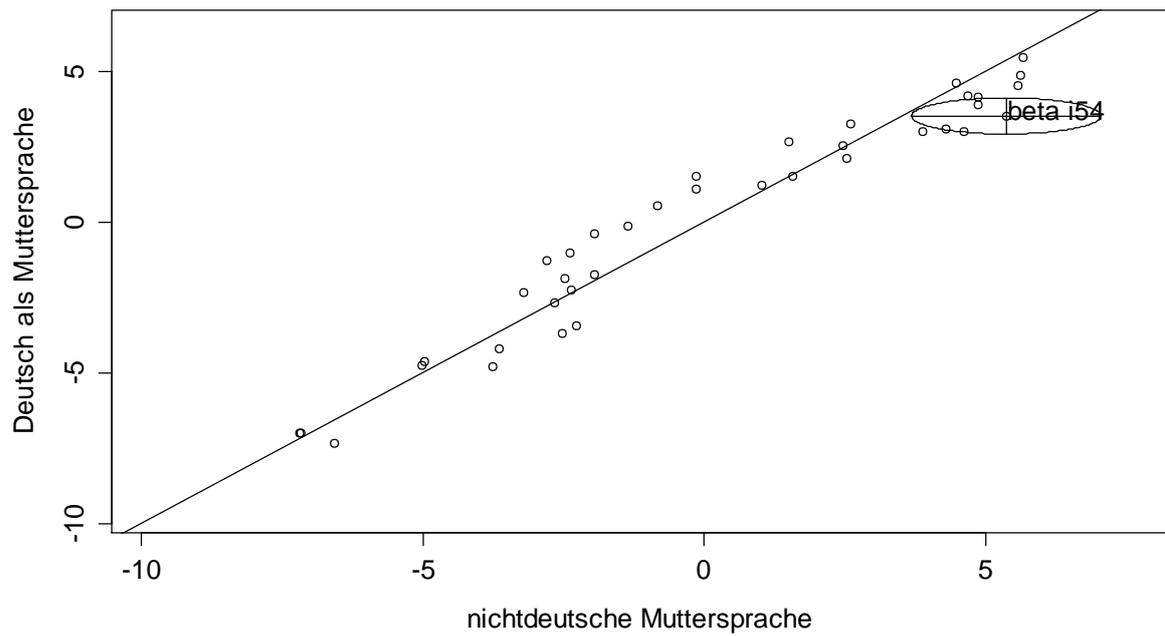
**Abbildung 5.13:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i48 ( $\alpha=0,01$ ); ( $n=443$ ).



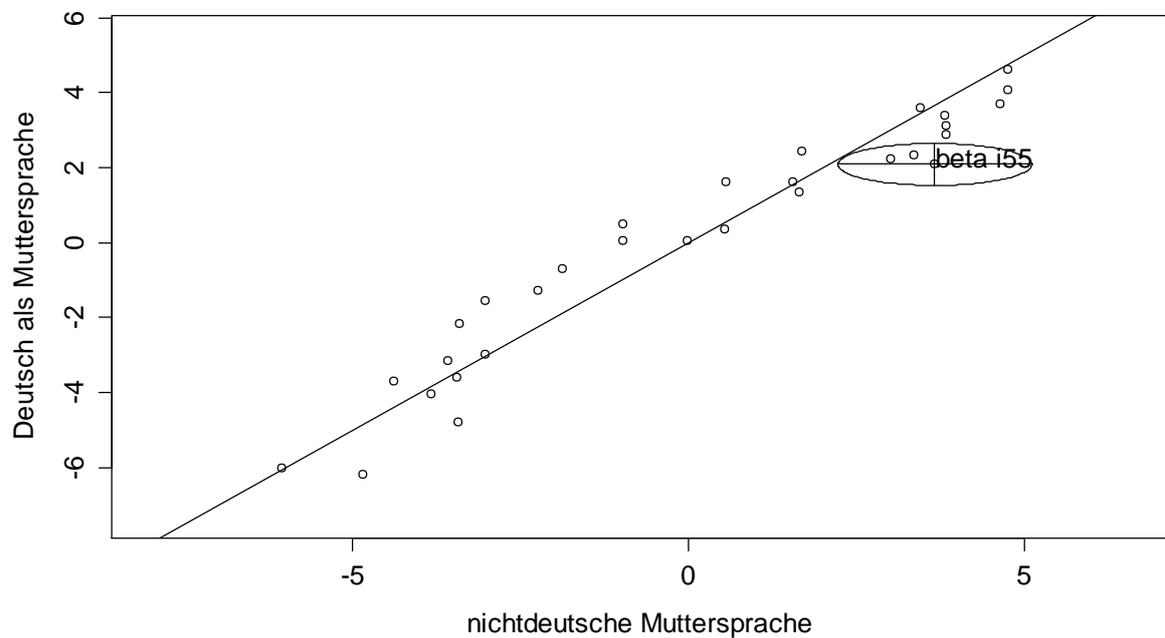
**Abbildung 5.14:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i50 ( $\alpha=0,01$ ); ( $n=443$ ).



**Abbildung 5.15:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i51 ( $\alpha=0,01$ ); ( $n=443$ ).



**Abbildung 5.16:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i54 ( $\alpha = 0,01$ ); ( $n=443$ ).



**Abbildung 5.17:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 1 mit Standardschätzfehler des Items i55 ( $\alpha = 0,01$ ); ( $n=443$ ).

Nach dem schrittweisen Ausschluss der 14 auffälligen Items kann dem UT 1 a-posteriori in der vorgenommenen retrospektiven Analyse für die erhobene Stichprobe Modellgültigkeit zugesprochen werden (siehe Tabelle 5.2).

**Tabelle 5.2:** Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 1 nach dem Ausschluss von sieben Aufgaben (n=443).

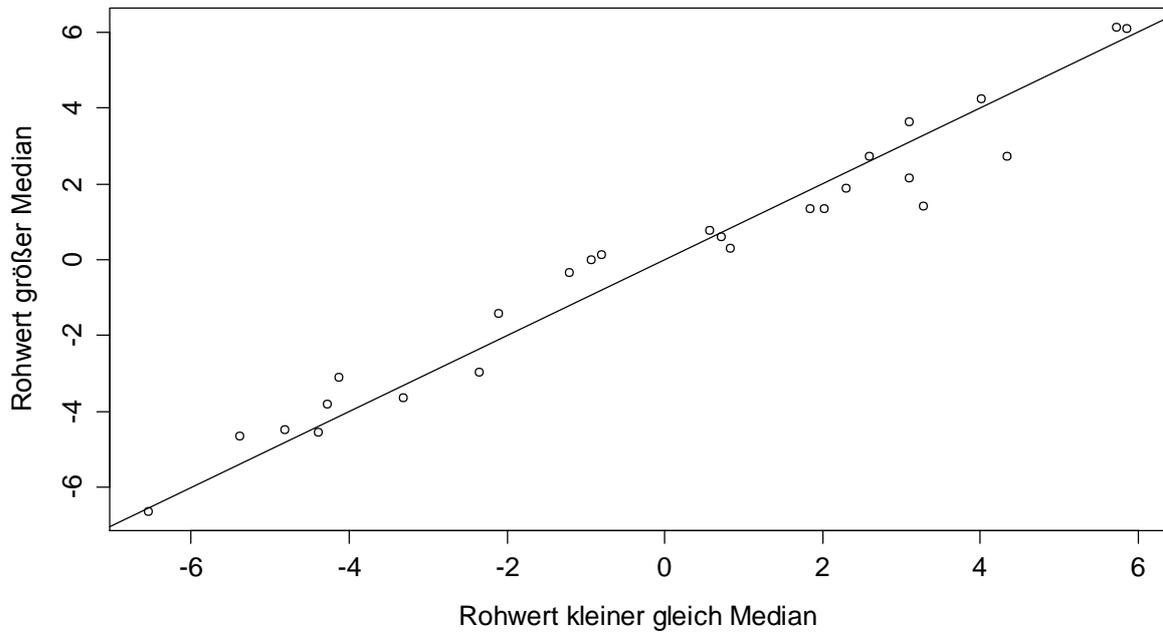
<b>Teilungskriterium</b>	$\chi^2$	<b>df</b>	$\chi^2_{\text{krit}} (\alpha=0,01)$	<b>p</b>
Score	36,113	25	44,314	0,070
Geschlecht	43,411	25	44,314	0,013
Muttersprache	33,772	23	41,638	0,069

Bei der Analyse von UT 1 fällt auf, dass nur die Aufgaben i54 und i55, welche für den AID 3 (Version 1.1) aus inhaltlichen Überlegungen ausgeschlossen wurden auch nach der Reanalyse des AID 2 (Version 2.1 und Version 2.2) teststatistisch auffällig waren. Bis auf die Items i54 und i57 wurden alle Aufgaben die in Folge der Analyse für die erhobene Stichprobe ausgeschlossen werden mussten im AID 3 (Version 3.1) aus inhaltlichen Gründen nicht ausgeschlossen (vgl. Tabelle 5.3).

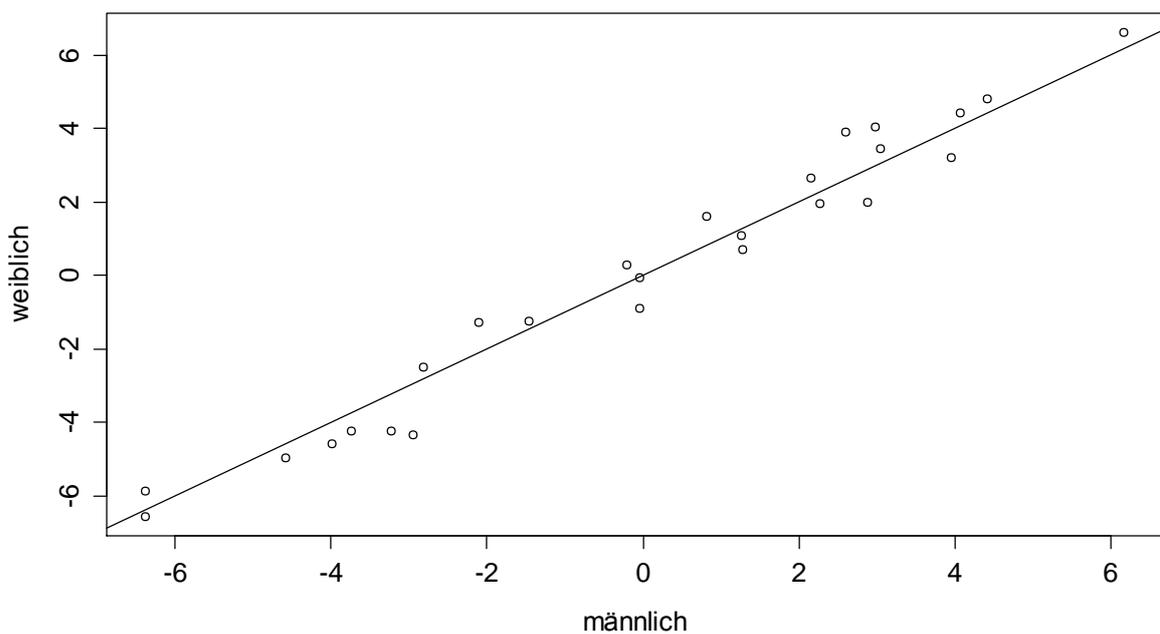
**Tabelle 5.3** Überblick zu den Items des Untertests 1 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.

<b>Untertest</b>	<b>Items die nicht in den AID3 übernommen wurden</b>	<b>Items die nicht analysiert werden konnten</b>	<b>Items die ausgeschlossen werden mussten</b>
UT 1	i1	i1	i12
	i4	i2	i13
	i8	i3	i16
	i11	i4	i20
	i22	i5	i23
	i24	i6	i26
	i28	i7	i31
	i29	i8	i42
	i32	i9	i48
	i34	i10	i50
	i35	i14	i51
	i36	i27	i54
	i37	i31	i55
	i39	i32	i57
	i40	i33	
	i42	i34	
	i47	i35	
	i49	i36	
	i50	i37	
	i54	i38	
i57	i39		
i58	i40		
i59	i46		
i60	i56d		

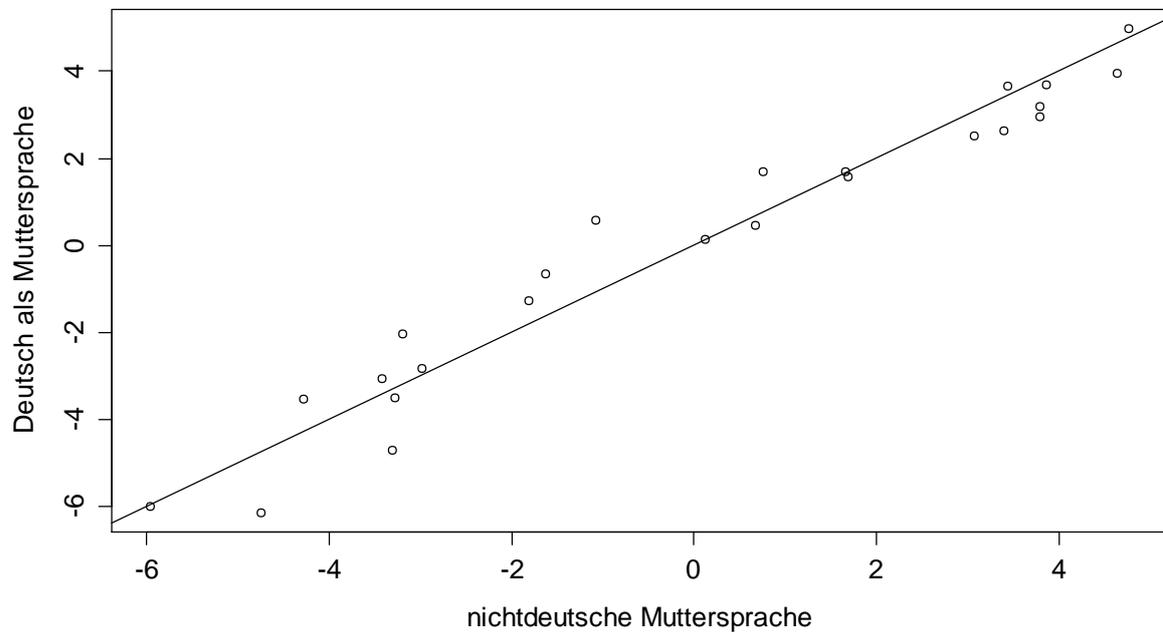
Nachfolgend werden die Grafischen Modellkontrollen der gewählten Teilungskriterien für den a-posteriori gültigen Itempool des Untertests 1 nach dem Ausscheiden von vierzehn Aufgaben in den Abbildungen 5.18 bis 5.20 dargestellt.



**Abbildung 5.18:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 1 nach dem Ausscheiden von vierzehn Items (n=443).



**Abbildung 5.19:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 1 nach dem Ausscheiden von vierzehn Items (n=443).



**Abbildung 5.20:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 1 nach dem Ausscheiden von vierzehn Items (n=443).

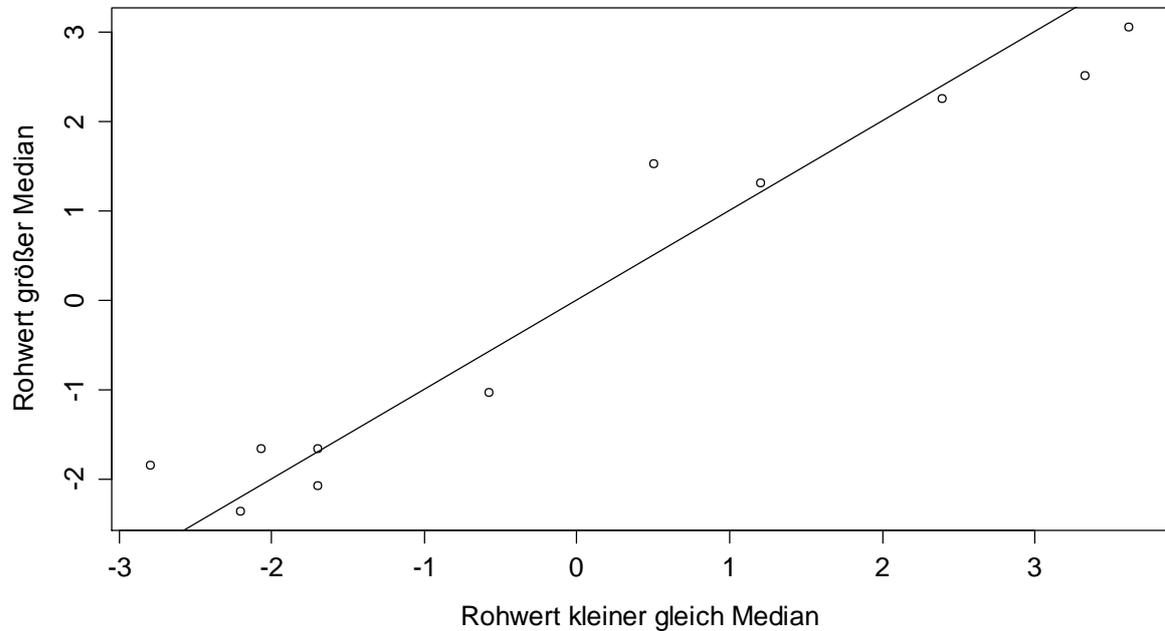
## 5.2 Untertest 2 Realitätssicherheit

Im Untertest 2 waren das Item i12 aufgrund der extremen Lösungshäufigkeiten (es wurde in der gesamten Stichprobe nur zwei Mal gelöst) und zusätzlich für das Teilungskriterium Score die Aufgabe i1 nicht analysierbar. Tabelle 5.4 gibt die signifikante Modellabweichung für das Tk Score vor dem Ausschluss dieser Items an.

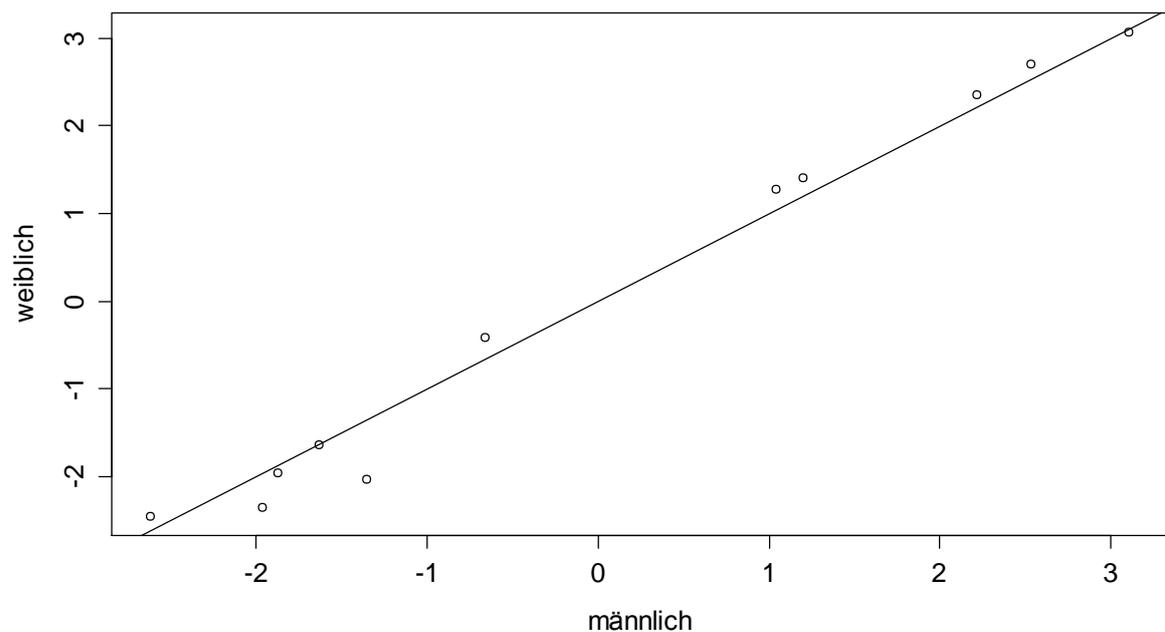
**Tabelle 5.4:** Darstellung der Ergebnisse der LQs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 2 vor dem Ausscheiden von Aufgaben (n=445).

Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	P
Score	33,387	10	23,209	<0,001
Geschlecht	8,320	10	23,209	0,598
Muttersprache	9,427	10	23,209	0,492

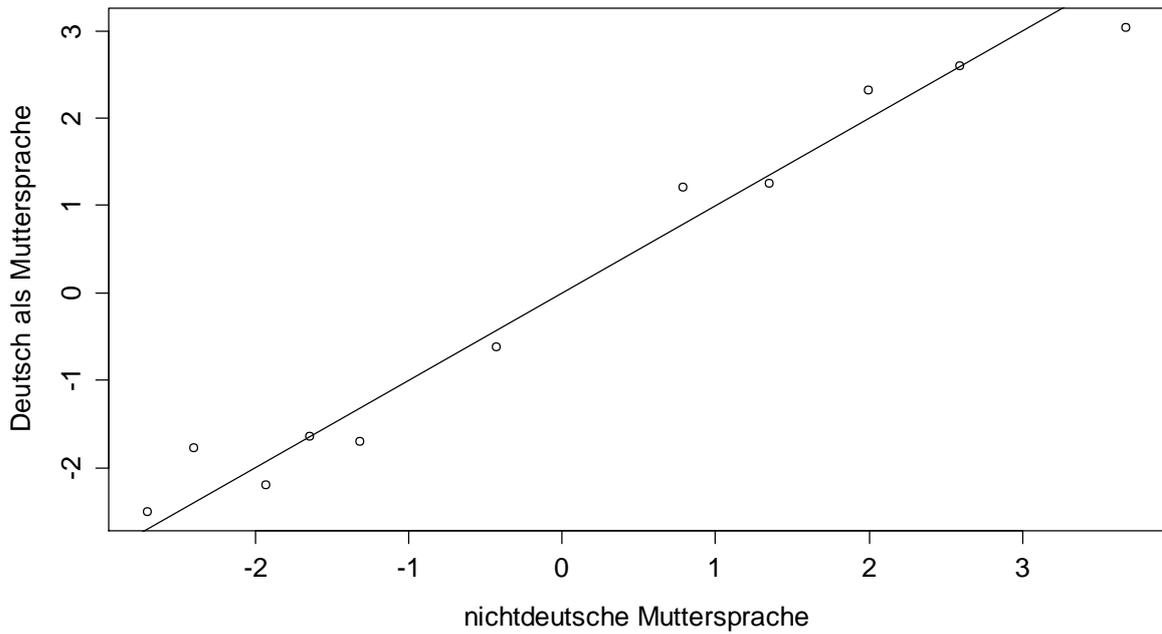
Die Abbildungen 5.21 bis 5.23 zeigen die grafischen Modellkontrollen der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 2 vor dem Ausschluss von Items.



**Abbildung 5.21:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 2 vor dem Ausschluss von Aufgaben (n=445).

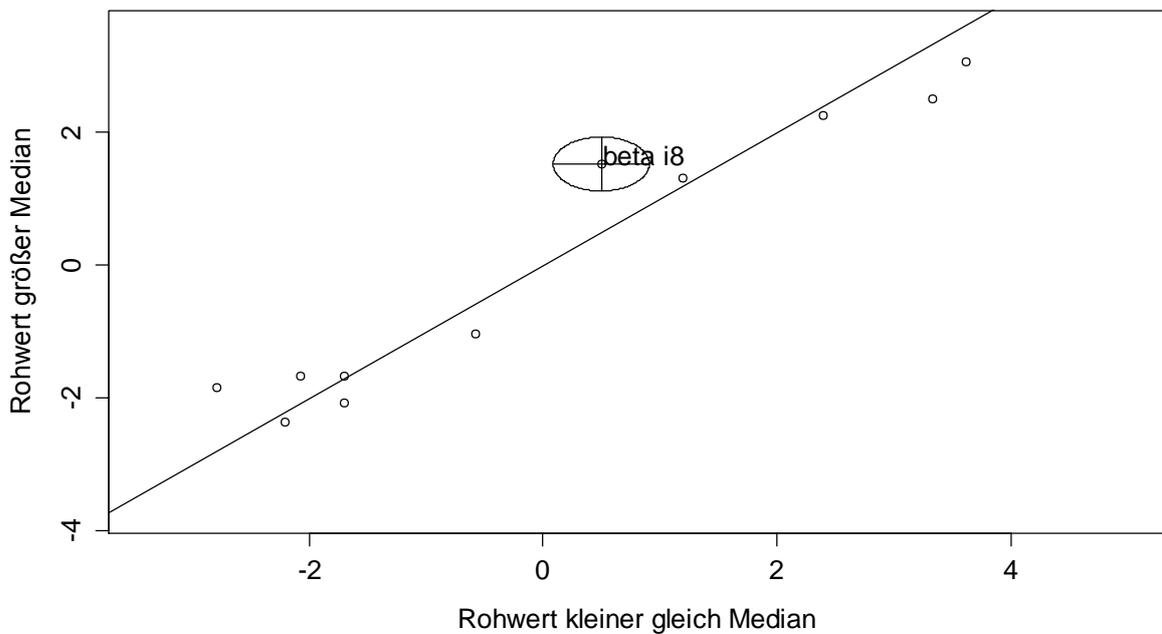


**Abbildung 5.22:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 2 vor dem Ausschluss von Aufgaben (n=445).



**Abbildung 5.23:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 2 vor dem Ausschluss von Aufgaben (n=445).

In der darauffolgenden Grafischen Modellkontrolle für das Tk Score war Item i8 auffällig (vgl. Abbildung 5.24).



**Abbildung 5.24:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 2 mit Standardschätzfehler des Items i8 ( $\alpha=0,01$ ); (n=445).

Nach dem Ausschluss der Aufgabe i8 konnte dem reduzierten Itempool des Untertests 2 für den AID 2 (Version 2.1 und Version 2.2) a-posteriori Modellgültigkeit zugeschrieben werden (vgl. Tabelle 5.5).

**Tabelle 5.5:** Darstellung der Ergebnisse der LQTs für den UT 2 für die Teilungskriterien Score, Geschlecht und Muttersprache nach dem Ausschluss einer Aufgabe (n=444).

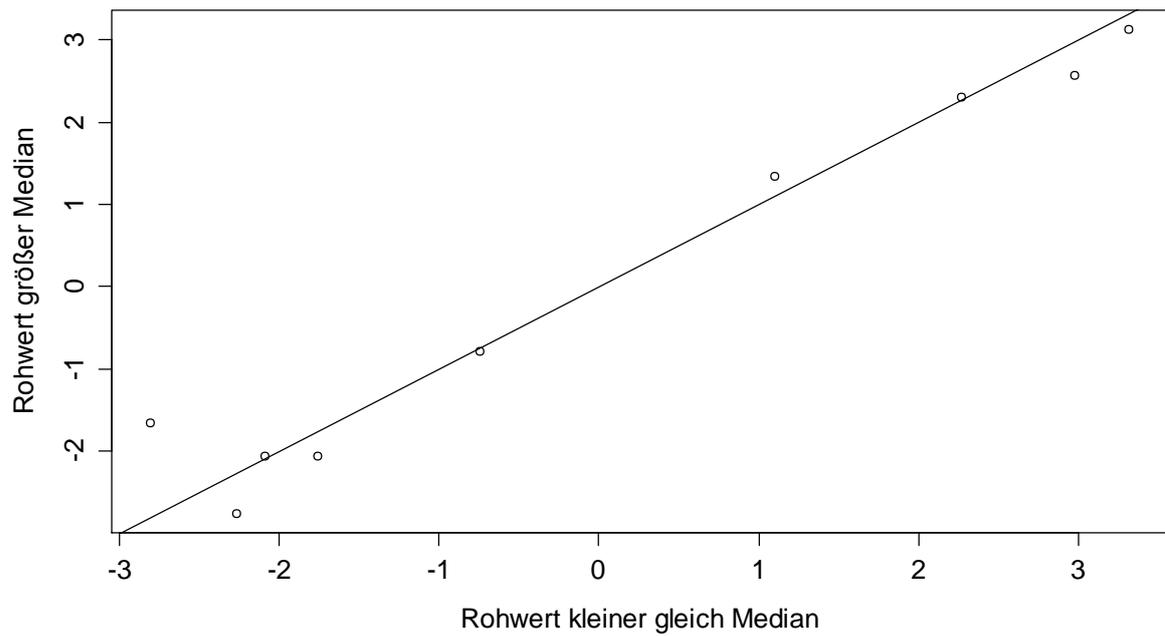
<b>Teilungskriterium</b>	$\chi^2$	<b>df</b>	$\chi^2_{\text{krit}} (\alpha=0,01)$	<b>p</b>
Score	5,367	8	20,090	0,718
Geschlecht	7,776	9	21,666	0,557
Muttersprache	6,986	9	21,666	0,639

Tabelle 5.6 zeigt jene Aufgaben des AID 2 (Version 2.1 und Version 2.2), welche auf Basis inhaltlicher Überlegungen nicht in den AID 3 (Version 3.1) übernommen wurden im Vergleich zu jenen Items die bei der vorgenommenen Analyse a) nicht analysierbar waren und b) jenes Items, welches aufgrund teststatistischer Berechnungen ausgeschlossen werden sollte.

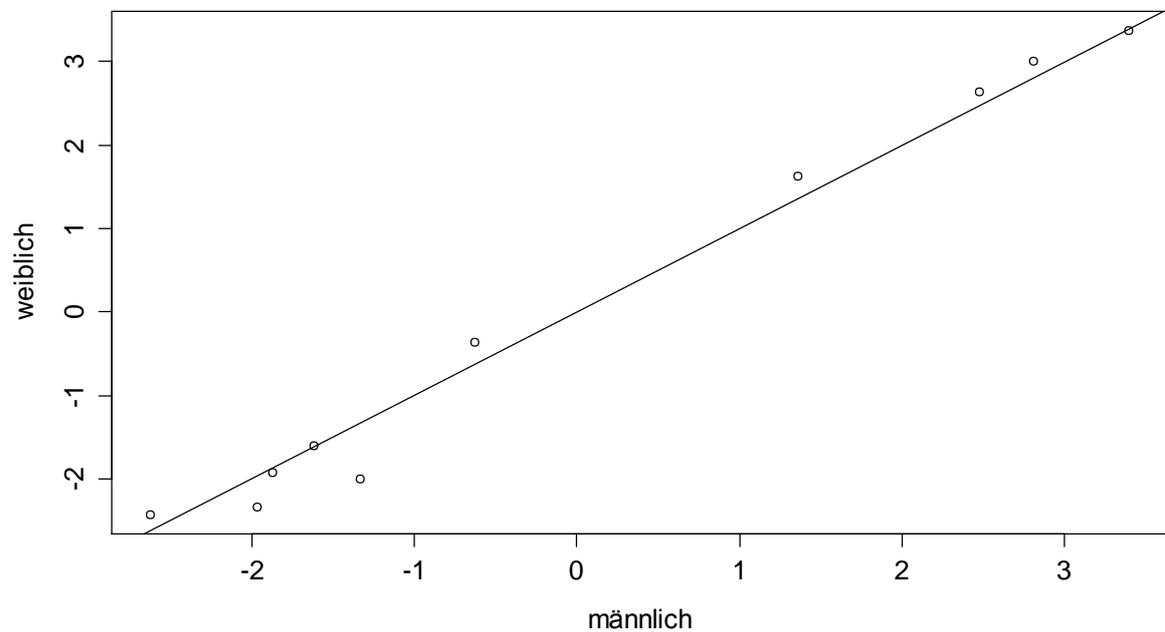
**Tabelle 5.6:** Überblick zu den Items des Untertests 2 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.

<b>Untertest</b>	<b>Items die nicht in den AID3 übernommen wurden</b>	<b>Items die nicht analysiert werden konnten</b>	<b>Items die ausgeschlossen werden mussten</b>
UT 2	i8 i9 i12	i12	i8

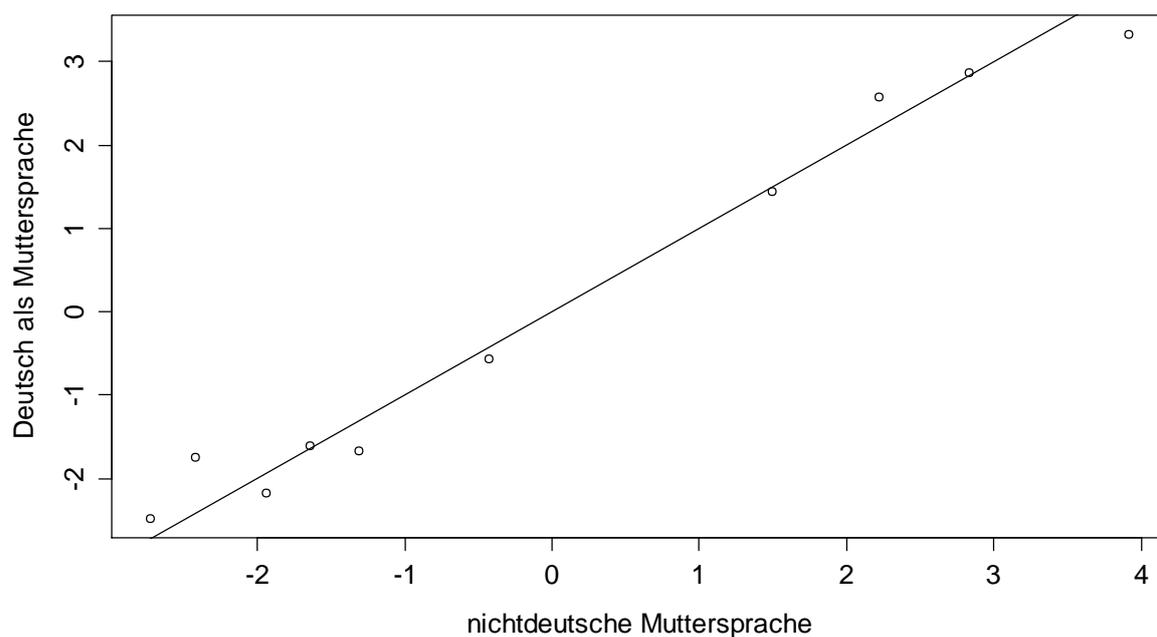
Die Abbildungen 5.25 bis 5.27 zeigen die Grafischen Modellkontrollen für den a-posteriori gültigen Itempool des Untertests 2.



**Abbildung 5.25:** Graphische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 2 nach dem Ausscheiden von einem Item (n=445).



**Abbildung 5.26:** Graphische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 2 nach dem Ausscheiden von einem Item (n=445).



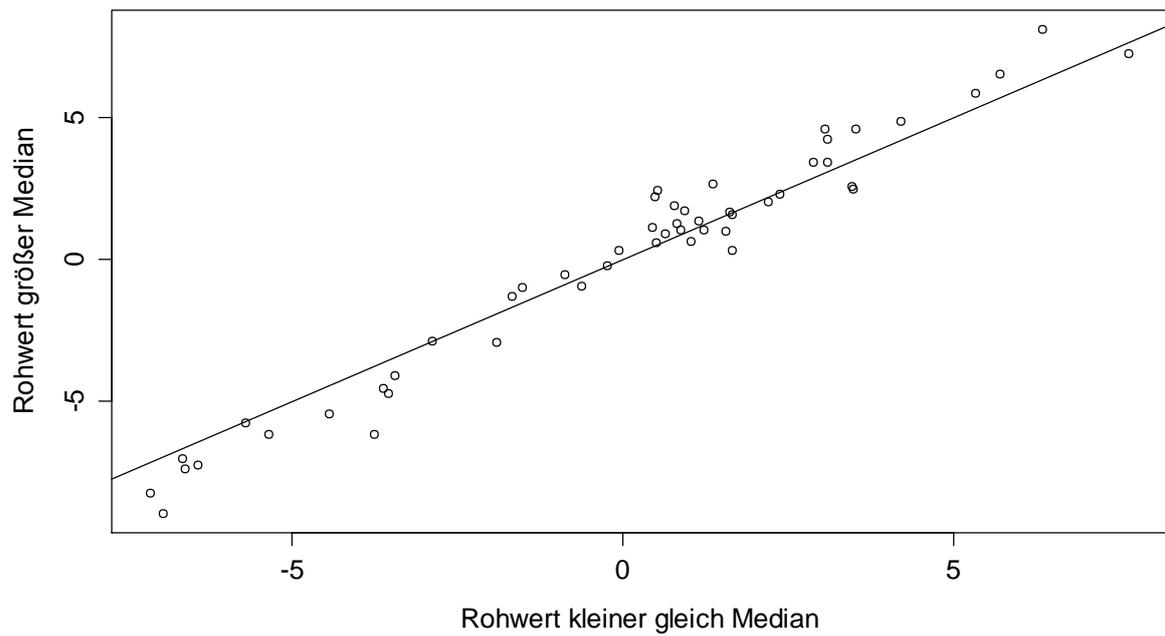
**Abbildung 5.27:** Graphische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 2 nach dem Ausscheiden von einem Item (n=445).

### 5.3 Untertest 3 Angewandtes Rechnen

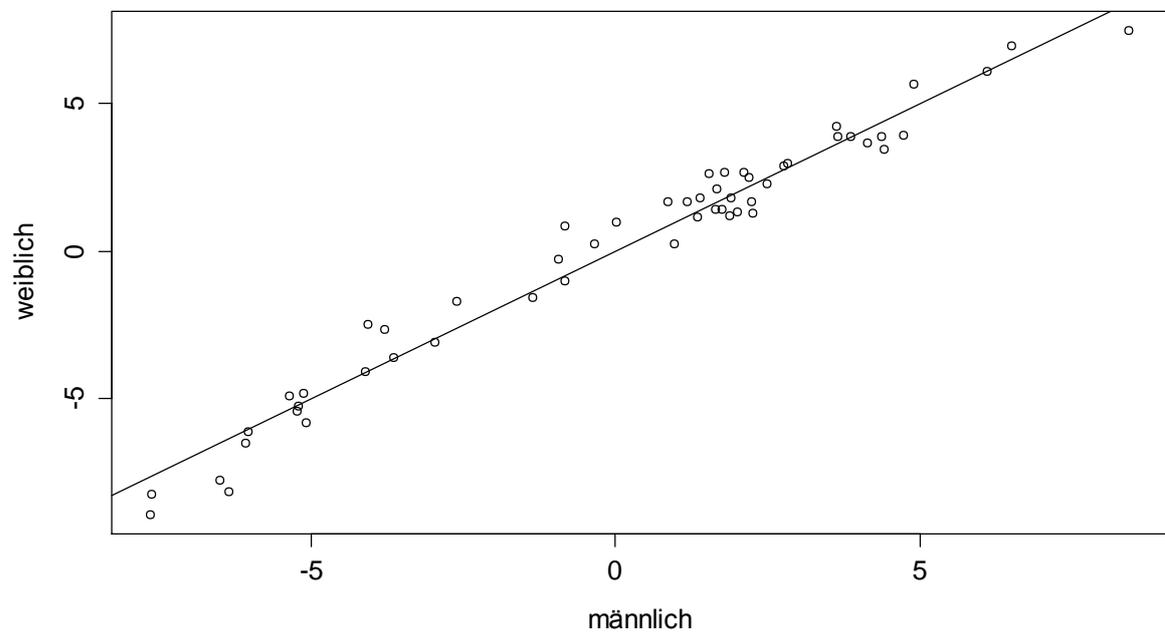
Im Untertest 3 musste Tp 0017 wegen fehlender Antworten ausgeschlossen werden. Die Items i1, i2, i3, i5, i8, i34, i37, i38, i39, i40, i41 und i43 waren nicht analysierbar. Tabelle 5.6 zeigt eine signifikante Modellabweichung vor dem Ausschluss des ersten Items. Die Abbildungen 5.28 bis 5.30 zeigen die Grafischen Modellkontrollen der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 3 vor dem Ausschluss von Items.

**Tabelle 5.7:** Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 3 vor dem Ausscheiden von Aufgaben (n=444).

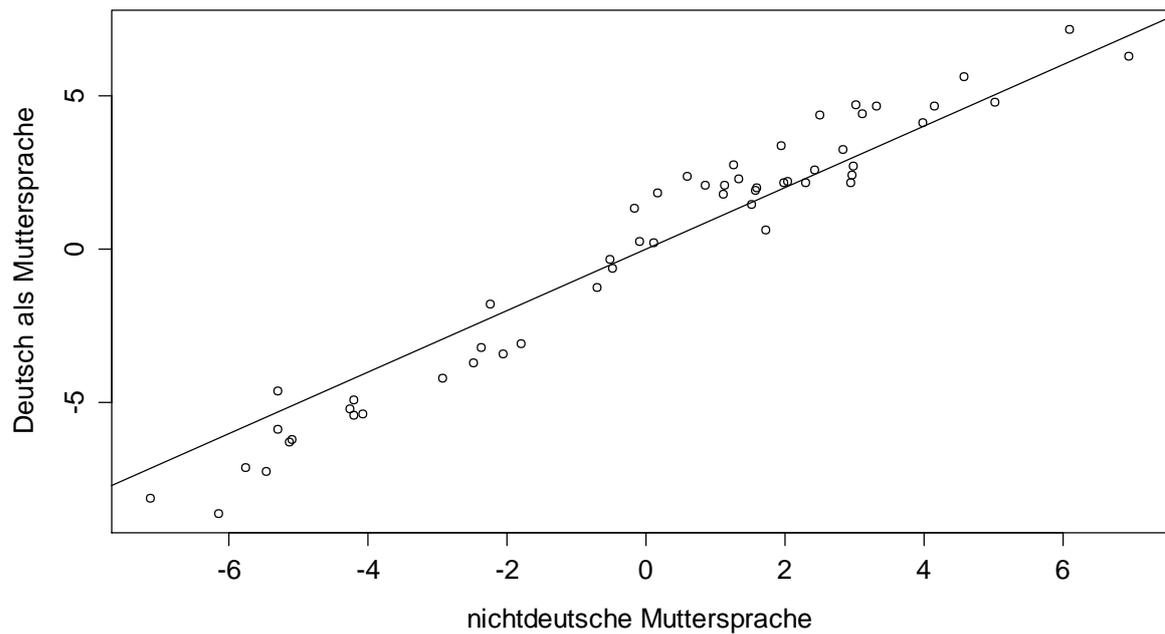
<b>Teilungskriterium</b>	$\chi^2$	<b>df</b>	$\chi^2_{\text{krit}} (\alpha=0,01)$	<b>P</b>
Score	89,278	50	76,154	0,001
Geschlecht	75,87	53	79,843	0,021
Muttersprache	75,474	53	79,843	0,023



**Abbildung 5.28:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 3 vor dem Ausschluss von Aufgaben (n=444).

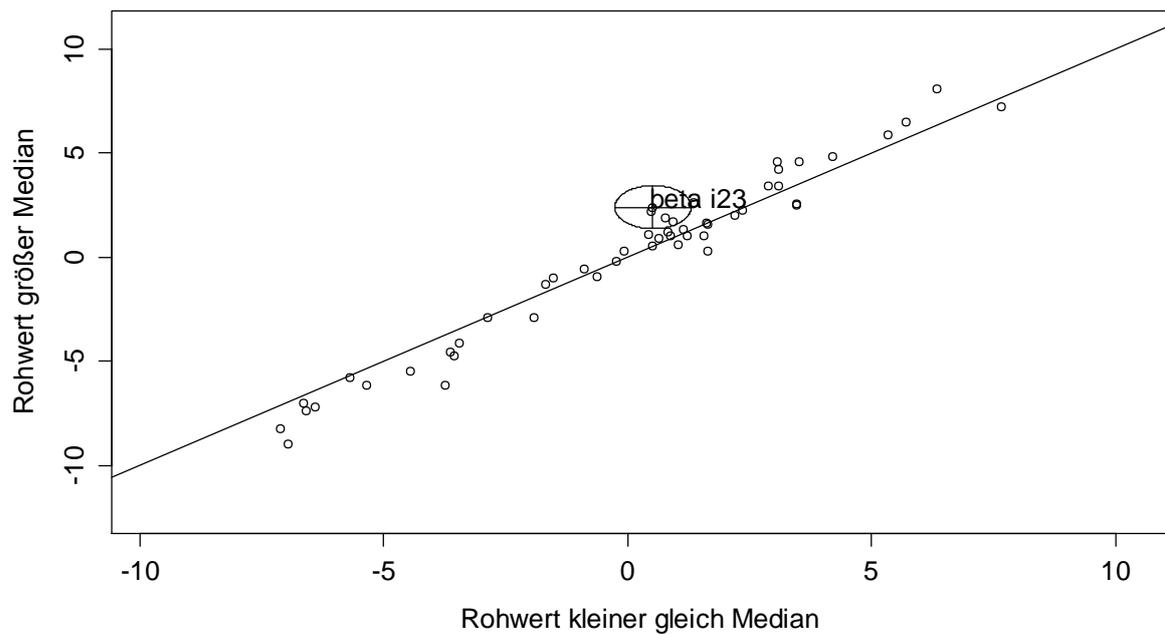


**Abbildung 5.29:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 3 vor dem Ausschluss von Aufgaben (n=444).

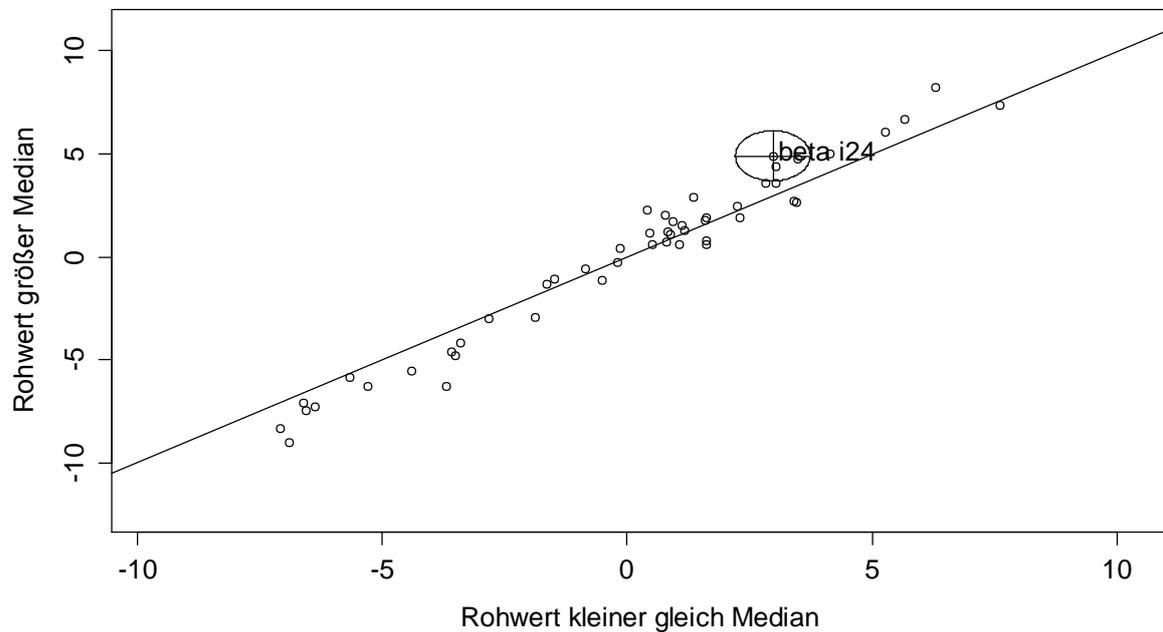


**Abbildung 5.30:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 3 vor dem Ausschluss von Aufgaben (n=444).

In den darauffolgenden Graphischen Modellkontrollen für das Tk Score waren die Aufgaben i23 und i24 auffällig (vgl. Abbildungen 5.31 und 5.32).



**Abbildung 5.31:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 3 mit Standardschätzfehler des Items 23 ( $\alpha=0,01$ ) ( $n=444$ ).

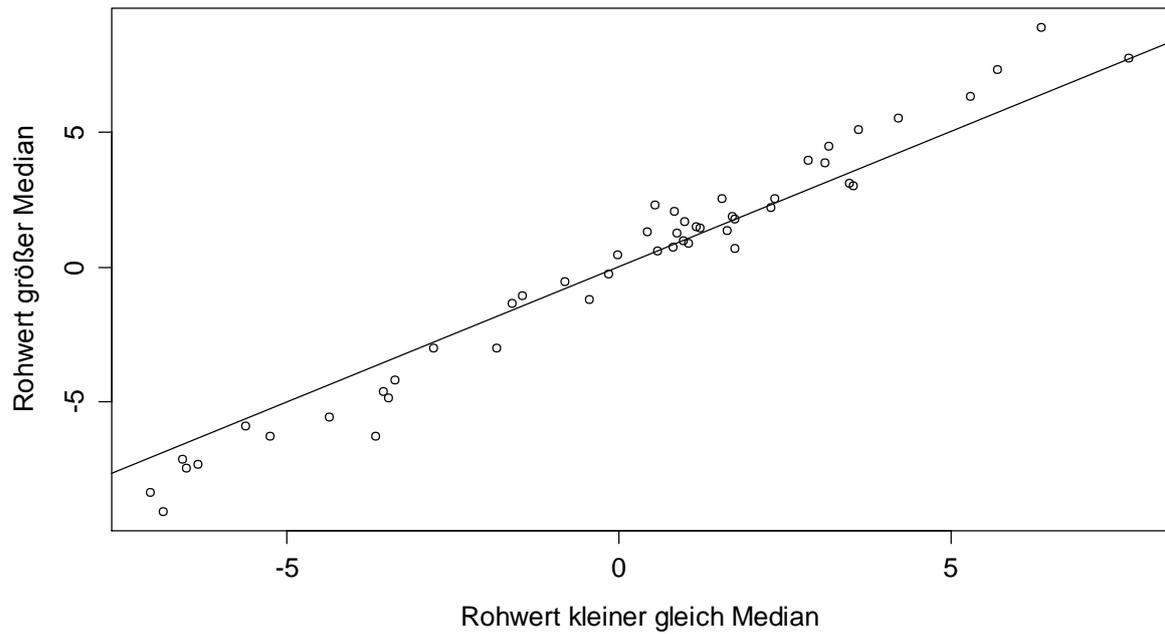


**Abbildung 5.32:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 3 mit Standardschätzfehler des Items 24 ( $\alpha=0,01$ ) ( $n= 444$ ).

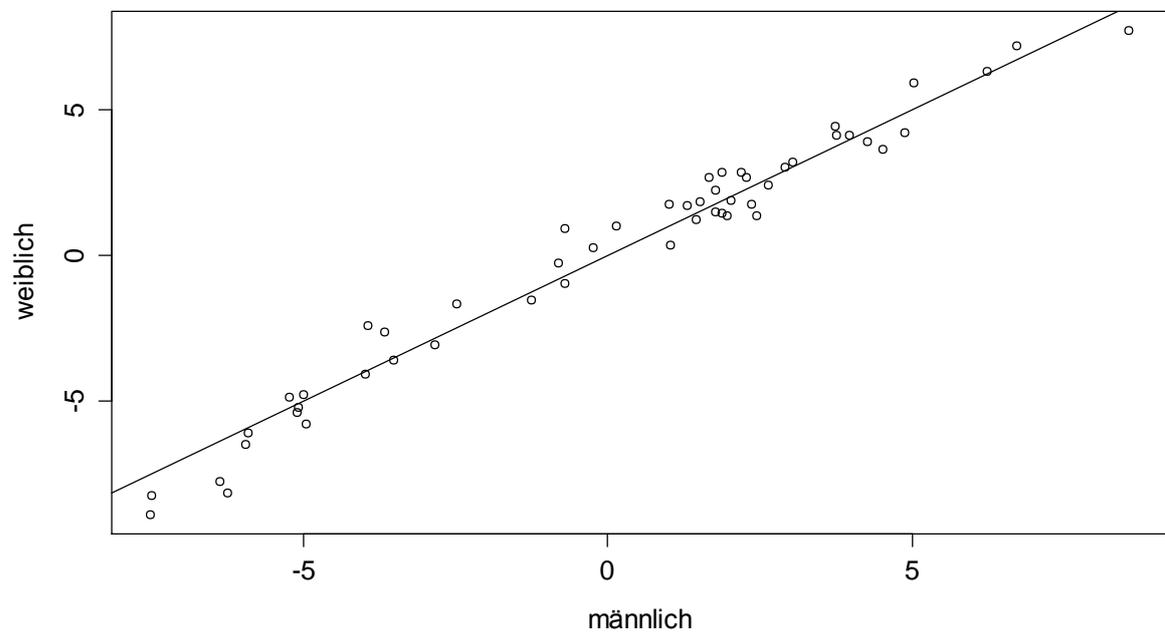
Nach dem schrittweisen Ausschluss der Aufgaben i23 (Range der Itemleichtigkeitsparameter von i23 im LQT mit -2,298 größer als ein Zehntel der Range aller Itemparameterschätzungen) und i24 (Range der Itemleichtigkeitsparameter von i24 im LQT mit 1,938 größer als ein Zehntel der Range aller Itemparameterschätzungen) konnte dem Itempool des UT 3 für den AID 2 (Version 2.1 und Version 2.2) a-posteriori Modellgültigkeit zugeschrieben werden (vgl. Tabelle 5.8 und Abbildungen 5.33 bis 5.35).

**Tabelle 5.8:** Darstellung der Ergebnisse der LQTs für den UT 3 für die Teilungskriterien Score, Geschlecht und Muttersprache nach dem Ausschluss von zwei Aufgaben ( $n=444$ )

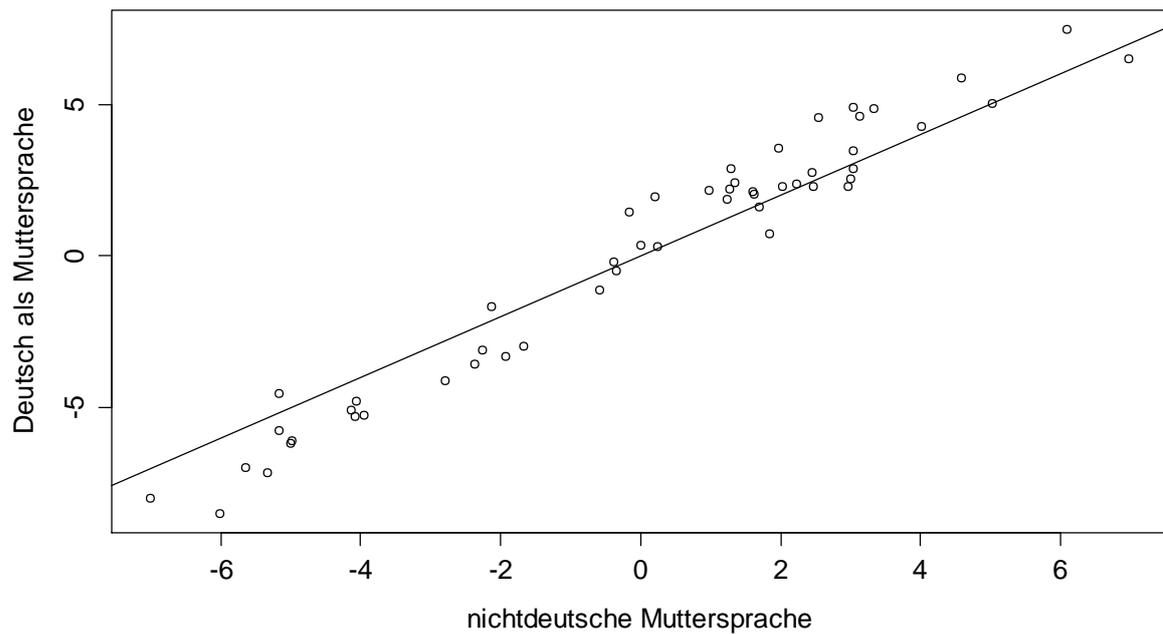
Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	p
Score	71,520	48	73,683	0,015
Geschlecht	72,461	51	77,386	0,026
Muttersprache	72,679	51	77,386	0,025



**Abbildung 5.33:** Graphische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 3 nach dem Ausscheiden von zwei Items.



**Abbildung 5.34:** Graphische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 3 nach dem Ausscheiden von zwei Items.



**Abbildung 5.35:** Graphische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 3 nach dem Ausscheiden von zwei Items.

Wie in Tabelle 5.9 ersichtlich ist somit der aus inhaltlichen Überlegungen vollzogene Ausschluss von Item i24 im AID 3 (Version 3.1) auch teststatistisch nachvollziehbar. Bei der in dieser Diplomarbeit untersuchten Stichprobe war auch Item i23 auffällig, welches im AID 3 (Version 3.1) beibehalten wurde.

**Tabelle 5.9:** Überblick zu den Items des Untertests 3 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.

<b>Untertest</b>	<b>Items die nicht in den AID3 übernommen wurden</b>	<b>Items die nicht analysiert werden konnten</b>	<b>Items die ausgeschlossen werden mussten</b>
UT 3	i2	i1	i23
	i6	i2	i24
	i24	i3	
	i26	i5	
	i36	i8	
	i37	i34	
	i39	i37	
	i40	i38	
	i41	i39	
	i48	i40	
i50	i41		
	i53	i43	

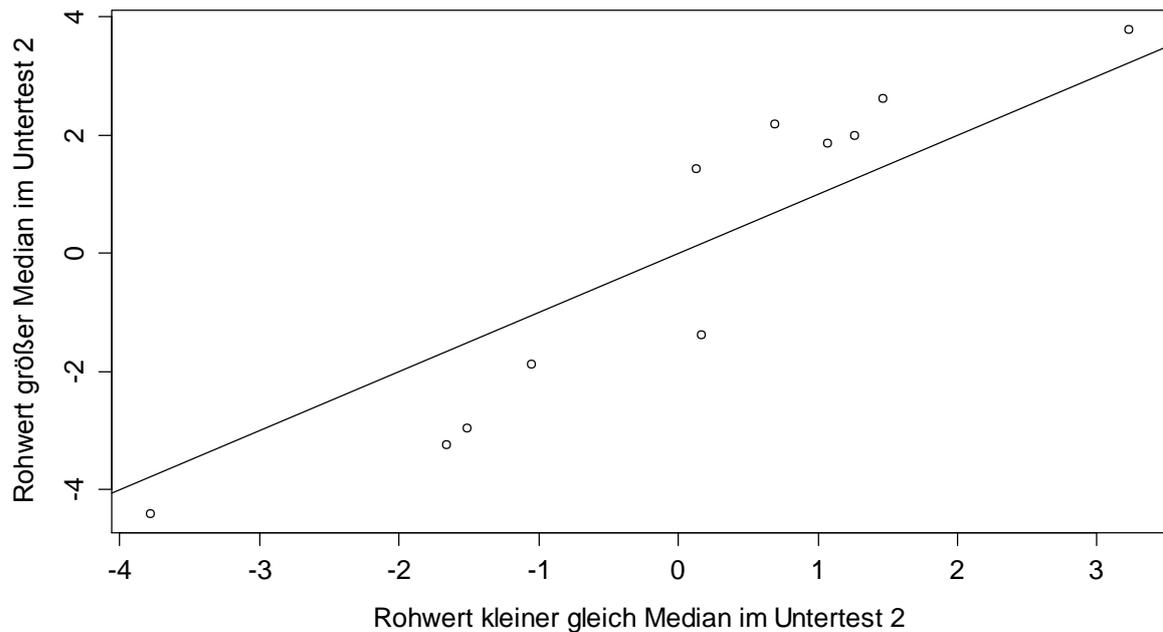
#### 5.4 Untertest 4 Soziale und Sachliche Folgerichtigkeit

Im Untertest 4 war in der erhobenen Stichprobe das Item i11 und nicht analysierbar. Für das Tk Score des Untertests 4 waren die Aufgaben i1, i2, i5, i6, i8, i11 und i17 nicht analysierbar. Aus Tabelle 5.10 ist das Modellkonforme Ergebnis für den UT 4 ohne Ausschluss eines Items ersichtlich.

**Tabelle 5.10:** Darstellung des Ergebnisses des LQTs für das Teilungskriterium Score im Untertest 4 vor dem Ausschluss von Aufgaben (n=445).

<b>Teilungskriterium</b>	$\chi^2$	<b>df</b>	$\chi^2_{\text{krit}} (\alpha=0,01)$	<b>p</b>
Score	11,553	10	23,210	0,316

Bei der Grafischen Modellkontrolle für das gewählte Teilungskriterium (Score im UT 2) waren keine Items auffällig. Es ist jedoch anzumerken, dass von den 18 Aufgaben in der erhobenen Stichprobe sieben nicht analysiert werden konnten (vgl. Abbildung 5.36).



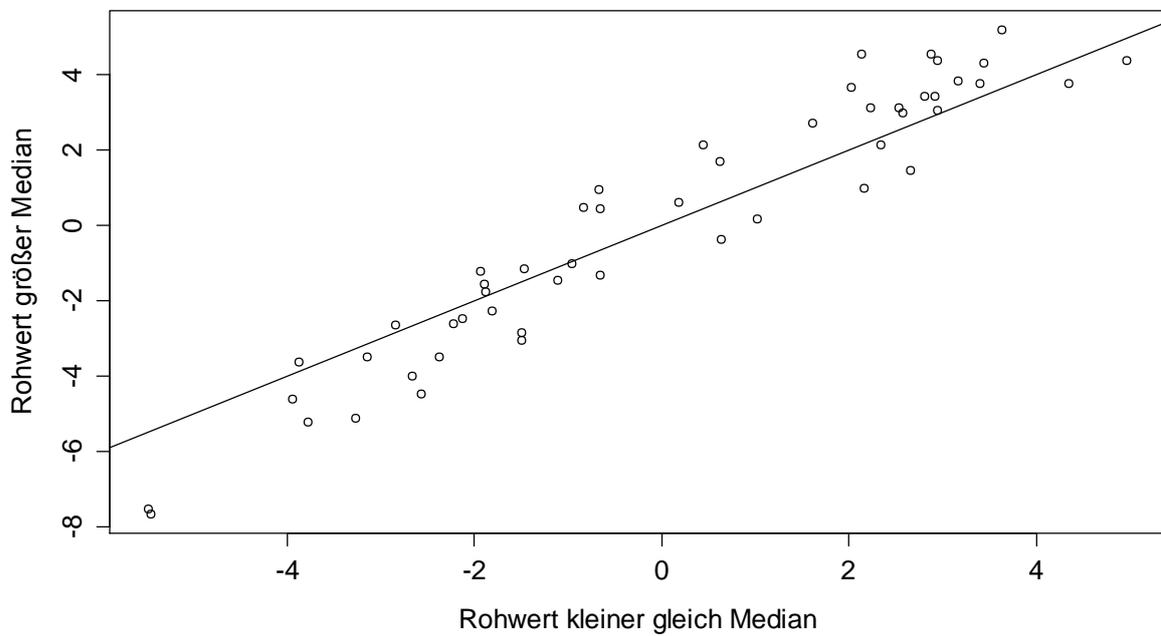
**Abbildung 5.36:** Grafische Modellkontrolle des Untertests 4 mittels des Teilungskriteriums Score im UT 2 (n=445).

## 5.5 Untertest 6 Synonyme Finden

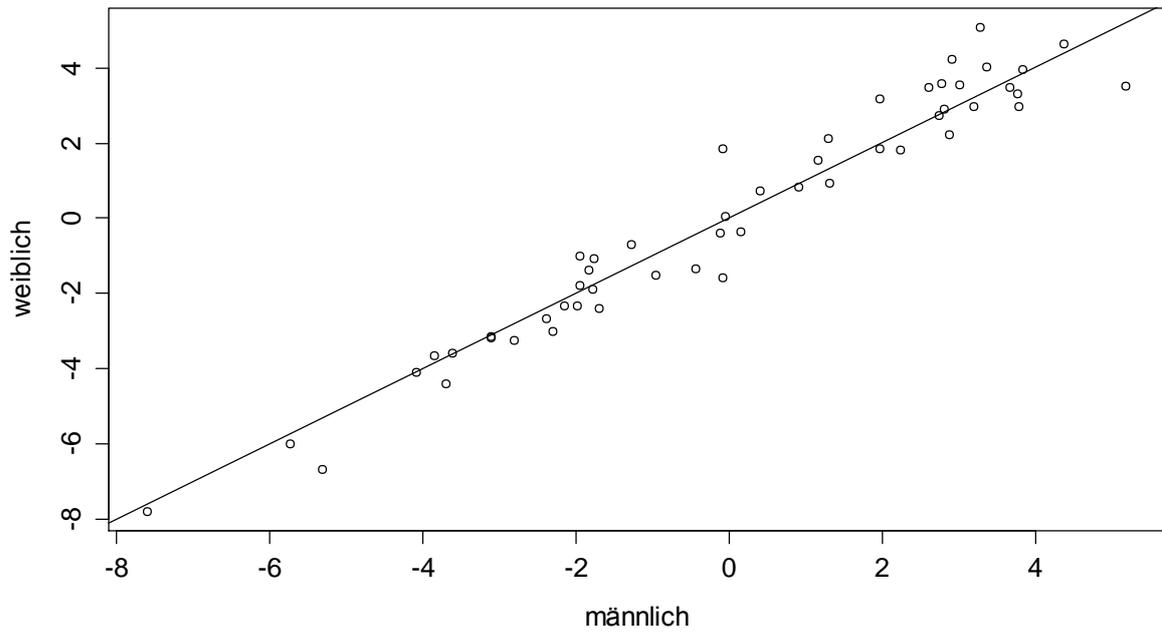
Im Untertest 6 waren die Aufgaben i1, i2, i3, i25, i28, i31, i32, i33, i34, i35, i36, i37, i38, i39, i40 und i59 in der erhobenen Stichprobe schlussendlich nicht analysierbar. Tabelle 5.11 zeigt signifikante Modellabweichungen für die erste Modellprüfung der Teilungskriterien Score und Muttersprache. Die Abbildungen 5.37 bis 5.39 zeigen die Grafischen Modellkontrollen der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 6 vor dem Ausschluss von Items.

**Tabelle 5.11:** Darstellung der Ergebnisse der LQs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 6 vor dem Ausschluss von Aufgaben (n=445).

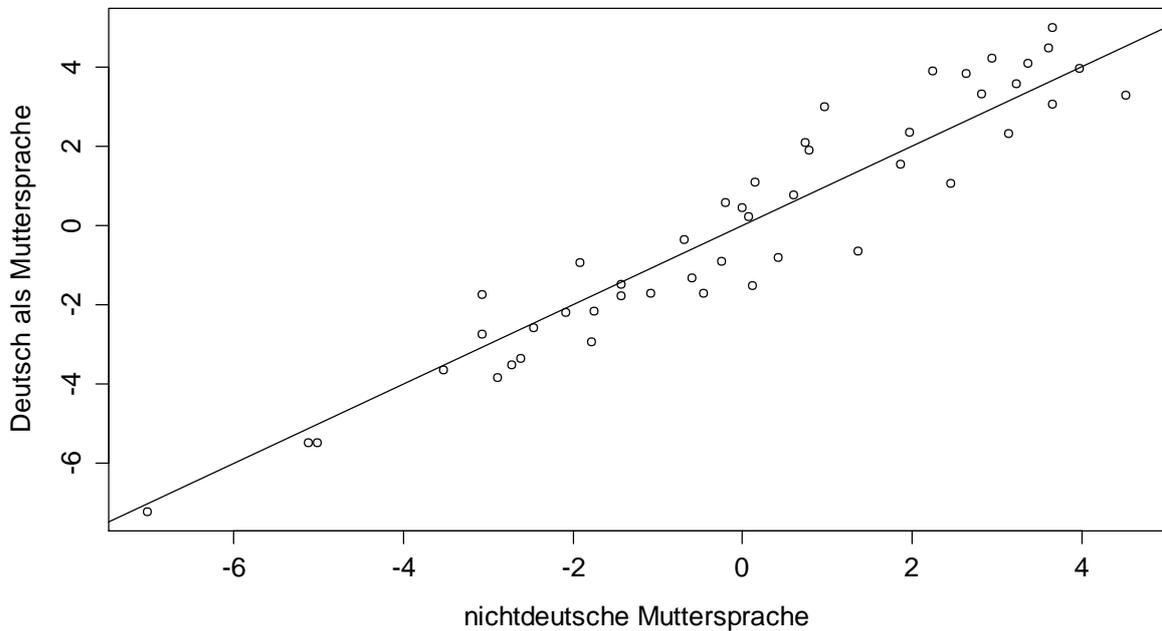
Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	P
Score	114,659	50	76,154	<0.001
Geschlecht	71,331	51	77,386	0,032
Muttersprache	78,946	46	71,201	0,002



**Abbildung 5.37:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 vor dem Ausschluss von Aufgaben (n=445).

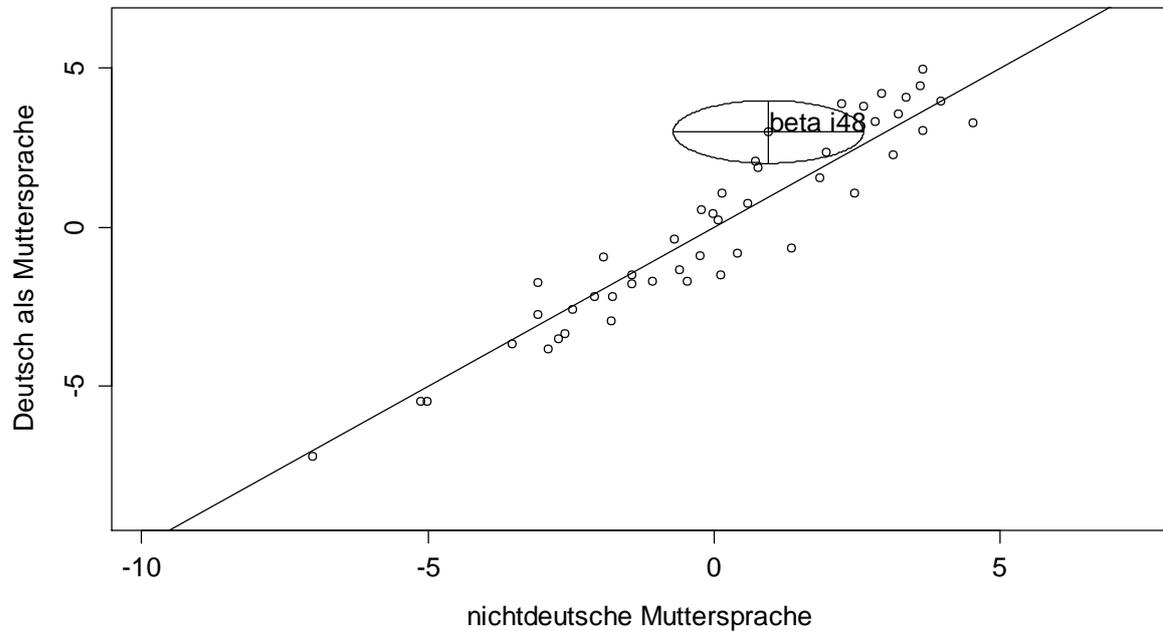


**Abbildung 5.38:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 6 vor dem Ausschluss von Aufgaben (n=445).



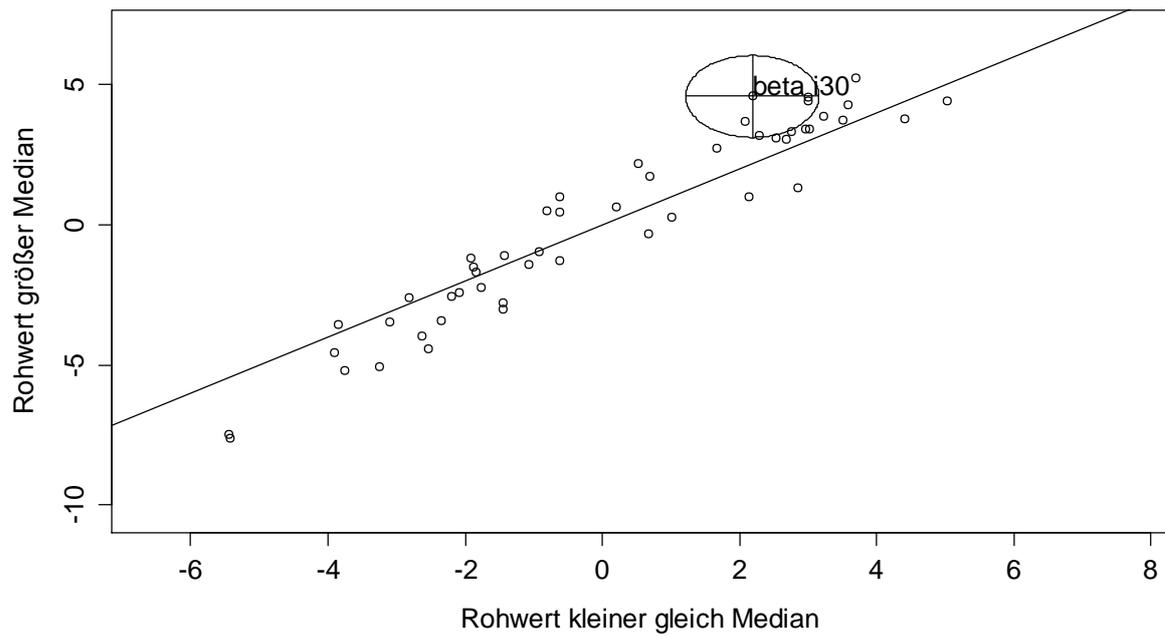
**Abbildung 5.39:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 6 vor dem Ausschluss von Aufgaben (n=445).

Bei der Grafischen Modellkontrolle für das Tk Muttersprache war Item i48 auffällig und wurde aufgrund der Differenz der Parameterschätzungen von 2,045 ausgeschlossen (vgl. Abbildung 5.40).

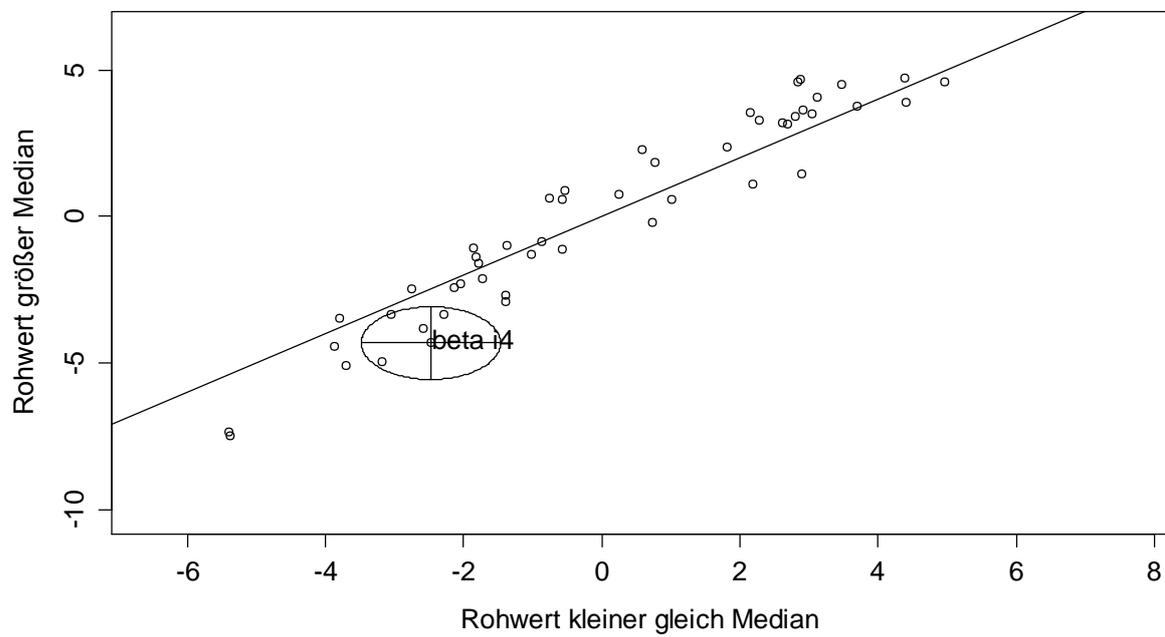


**Abbildung 5.40:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i48 ( $\alpha=0,01$ ); ( $n=445$ ).

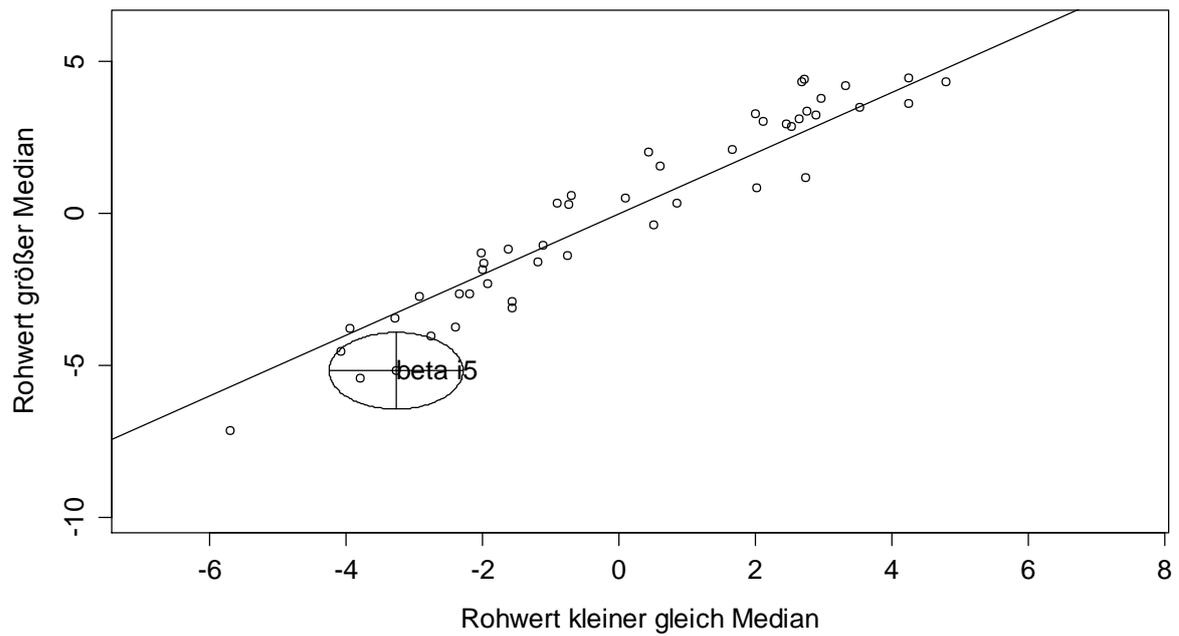
Bei der Grafischen Modellkontrolle für das Teilungskriterium Score waren die Items i30, i4, i5, i12, i10 und i23 auffällig. Diese wurden schrittweise ausgeschlossen (vgl. Abbildung 5.41 bis 5.46).



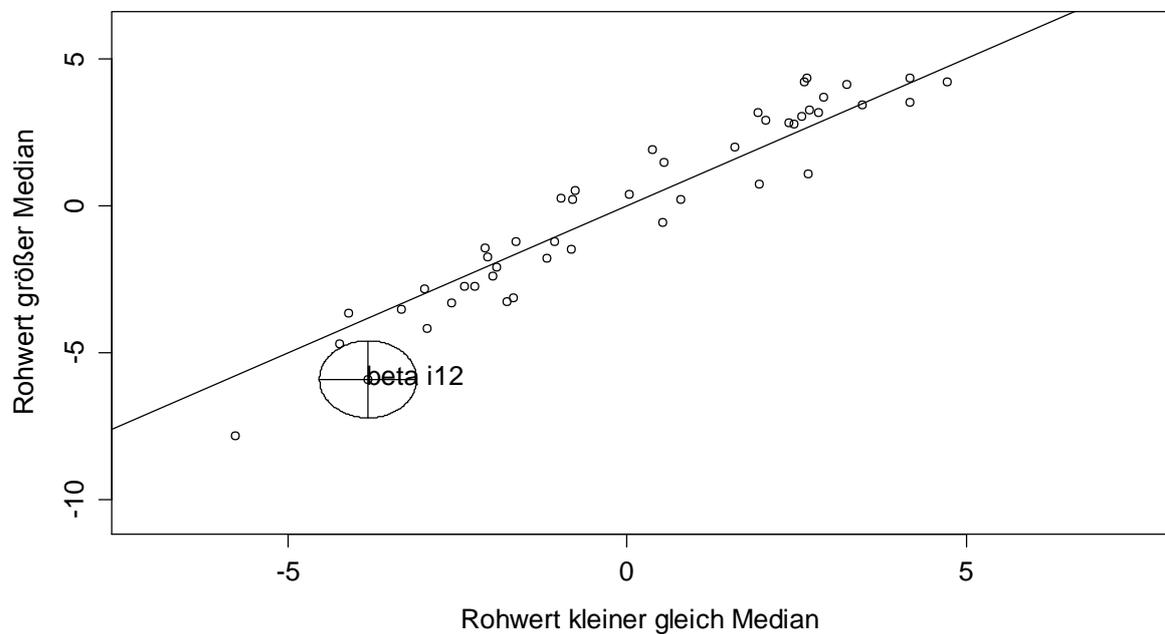
**Abbildung 5.41:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items  $i_{30} (\alpha=0,01)$ ; ( $n=445$ ).



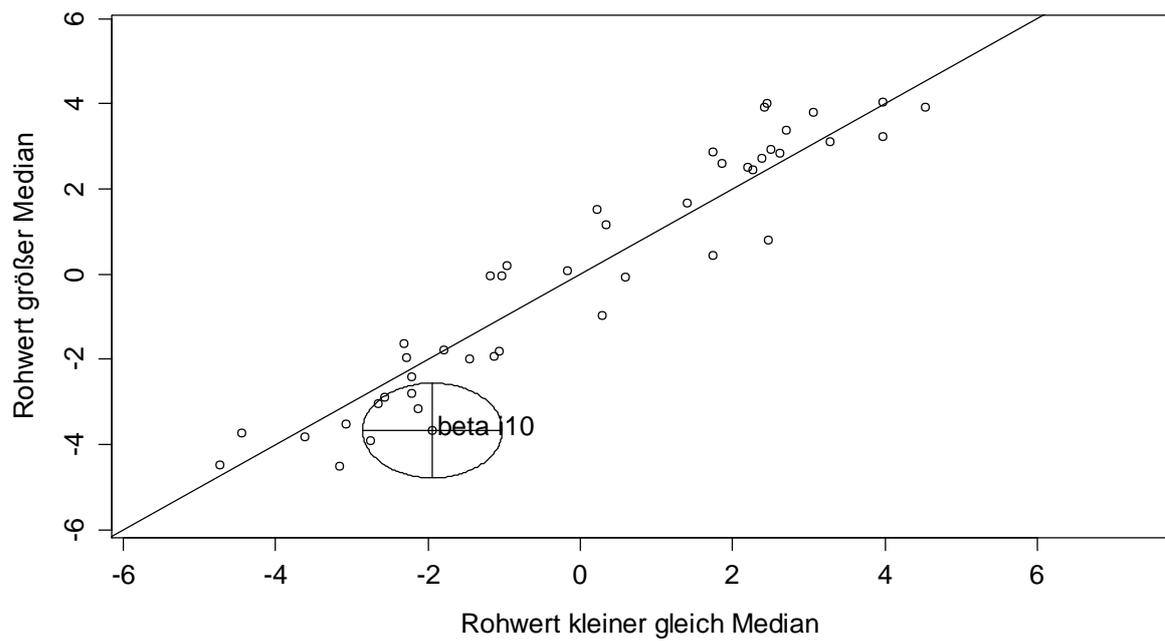
**Abbildung 5.42:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items  $i_4 (\alpha=0,01)$ ; ( $n=445$ ).



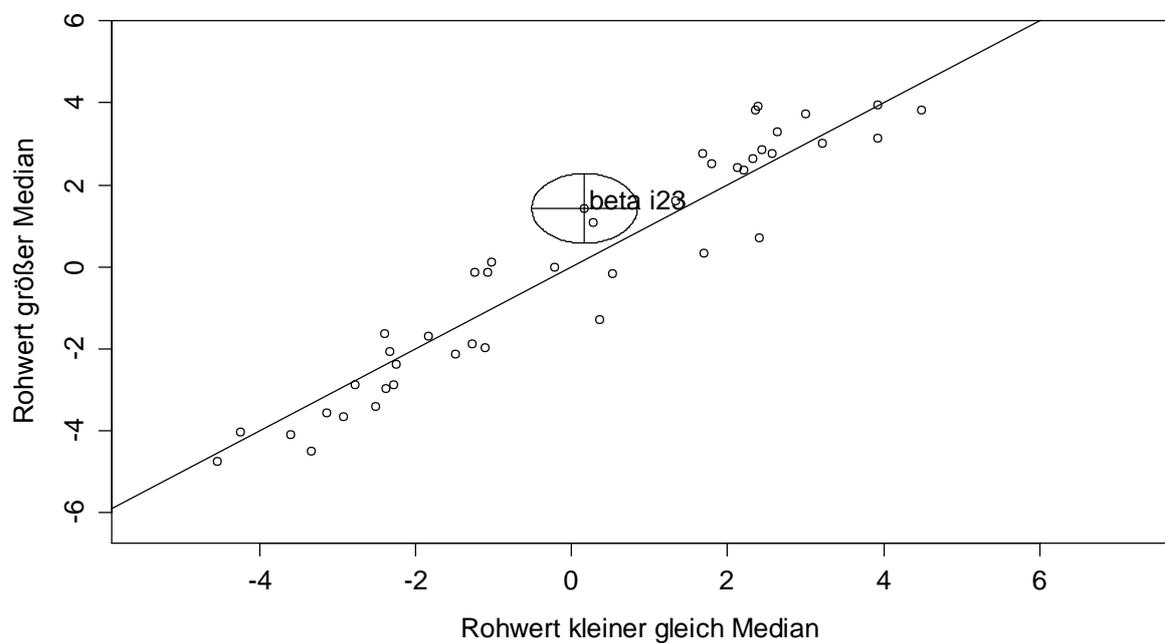
**Abbildung 5.43:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i5 ( $\alpha=0,01$ ); ( $n=445$ ).



**Abbildung 5.44:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i12 ( $\alpha=0,01$ ); ( $n=445$ ).



**Abbildung 5.45:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i10 ( $\alpha=0,01$ ); ( $n=445$ ).



**Abbildung 5.46:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 6 mit Standardschätzfehler des Items i23 ( $\alpha=0,01$ ); ( $n=445$ ).

Nach dem schrittweisen Ausschluss der sieben auffälligen Items kann dem UT 6 a-posteriori in der vorgenommenen Reanalyse für die erhobene Stichprobe Modellgültigkeit zugesprochen werden (siehe Tabelle 5.12).

**Tabelle 5.12:** Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 6 nach dem Ausschluss von sieben Aufgaben (n=445).

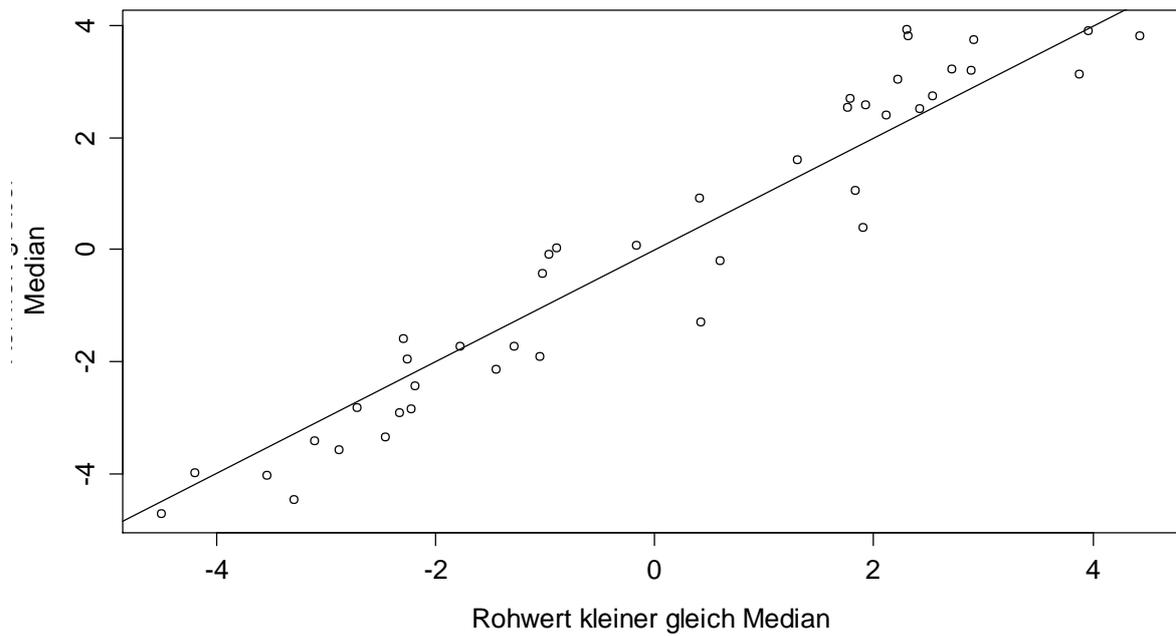
<b>Teilungskriterium</b>	$\chi^2$	<b>df</b>	$\chi^2_{\text{krit}} (\alpha=0,01)$	<b>p</b>
Score	57,232	41	64,950	0,047
Geschlecht	67,414	44	68,710	0,013
Muttersprache	61,865	39	62,428	0,011

Die Aufgaben i4, i5, i10, i12, i30 und i48 wurden für den AID 3 (Version 3.1) aus inhaltlichen Überlegungen ausgeschlossen. Diese Items waren nach der Reanalyse des AID 2 (Version 2.1 und Version 2.2) auch teststatistisch auffällig. Zusätzlich war auch das Item i23 auffällig und für die erhobene Stichprobe auszuschließen (vgl. Tabelle 5.13).

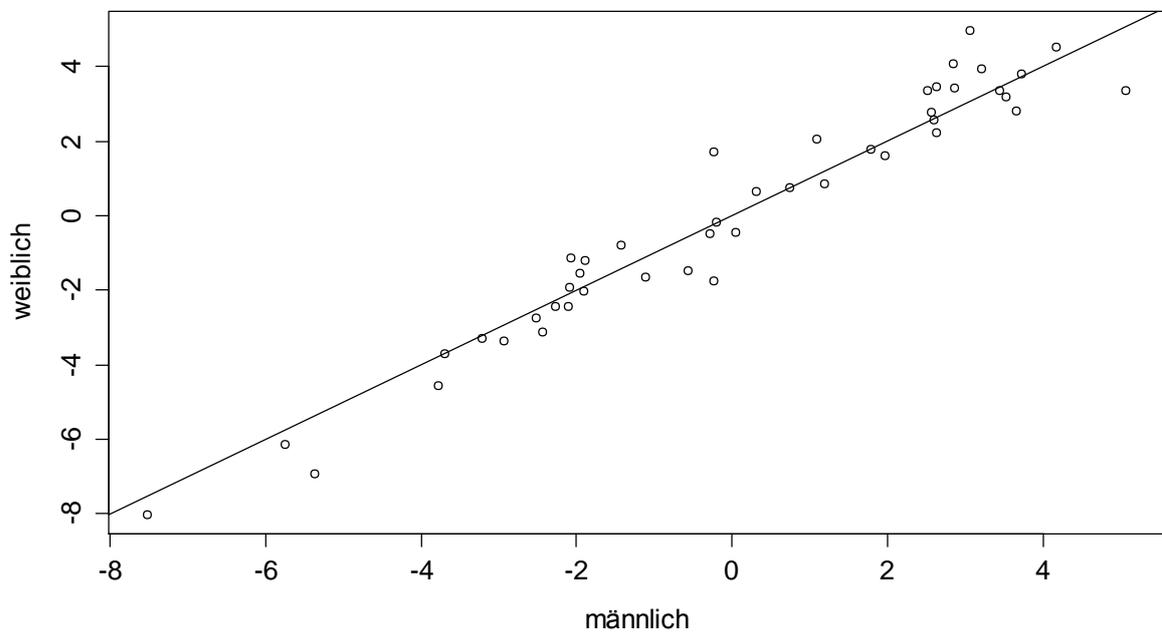
**Tabelle 5.13:** Überblick zu den Items des Untertests 6 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.

<b>Untertest</b>	<b>Items die nicht in den AID3 übernommen wurden</b>	<b>Items die nicht analysiert werden konnten</b>	<b>Items die ausgeschlossen werden mussten</b>
UT 6	i4	i34	i1
	i5	i35	i2
	i9	i36	i3
	i10	i37	i25
	i11	i39	i28
	i12	i40	i31
	i13	i41	i32
	i14	i46	i33
	i20	i47	i34
	i21	i48	i35
	i22	i49	i36
	i24	i50	i37
	i25	i51	i39
	i26	i53	i40
	i29	i57	i59
	i30	i59	
	i31		
i32			
i33			

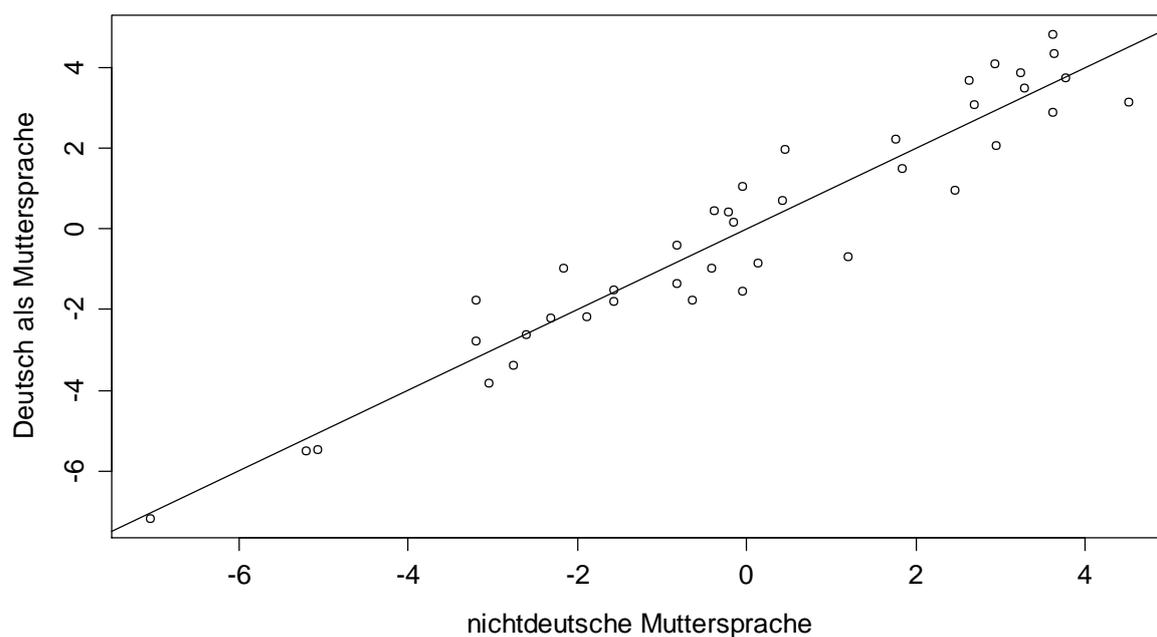
Die Grafischen Modellkontrollen der gewählten Teilungskriterien für den a-posteriori gültigen Itempool des Untertests 6 nach dem Ausscheiden von sieben Aufgaben können den Abbildungen 5.47 bis 5.49 entnommen werden.



**Abbildung 5.47:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 6 nach dem Ausscheiden von fünf Items (n=445).



**Abbildung 5.48:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 6 nach dem Ausscheiden von fünf Items (n=445).



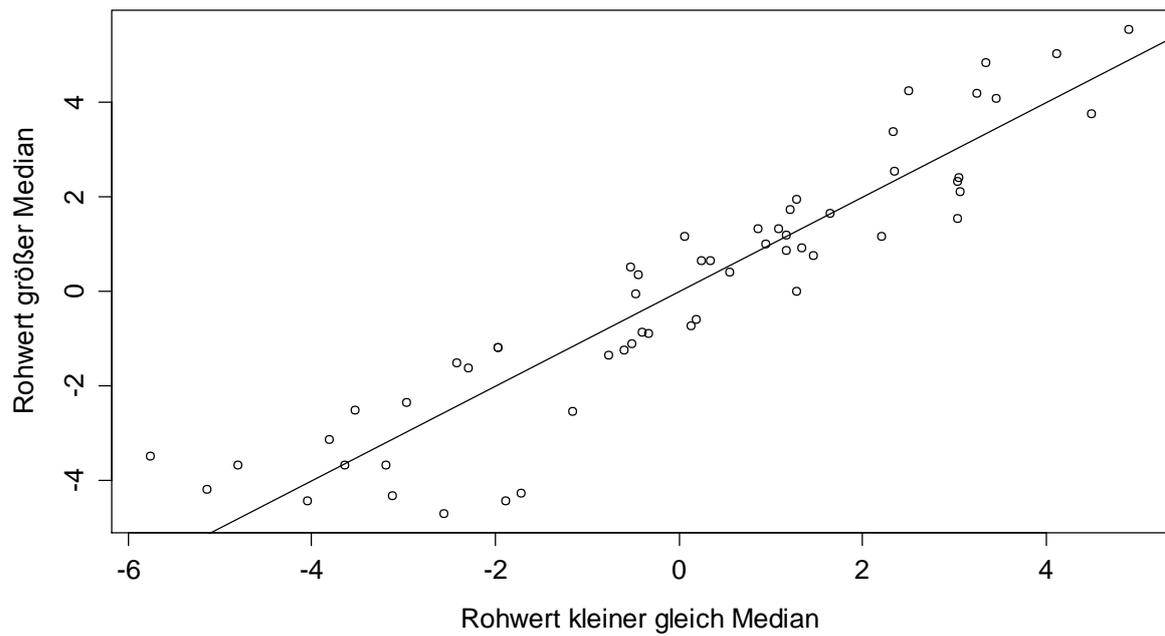
**Abbildung 5.49:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 6 nach dem Ausscheiden von fünf Items (n=445).

## 5.6 Untertest 9 Funktionen Abstrahieren

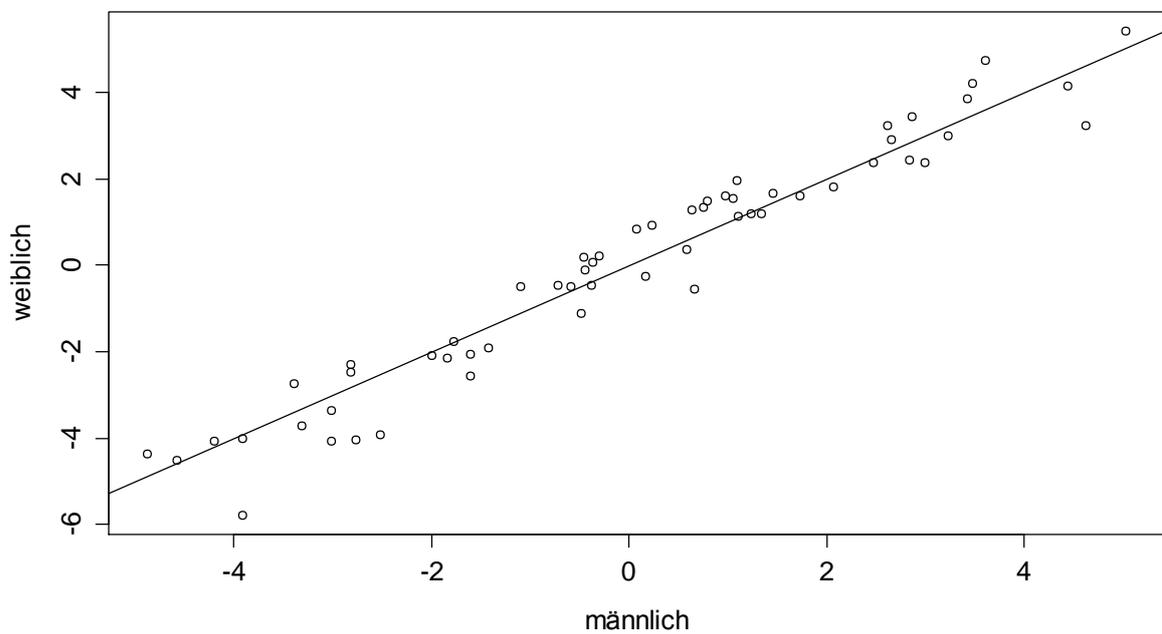
Im Untertest 9 mussten Tpn 0017, 0224 und 0340 wegen fehlender Antworten ausgeschlossen werden. Die Aufgaben i1, i3, i4 i37, i39, i44 und i48 waren in der erhobenen Stichprobe schlussendlich nicht analysierbar. Aus Tabelle 5.14 ist ersichtlich dass die erste Modellprüfung des Untertests 9 eine signifikante Modellabweichung für das Tk Score ergab. Die Abbildungen 5.50 bis 5.52 zeigen die Grafischen Modellkontrollen der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 9 vor dem Ausschluss von Items.

**Tabelle 5.14:** Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 9 vor dem Ausschluss von Aufgaben (n=442).

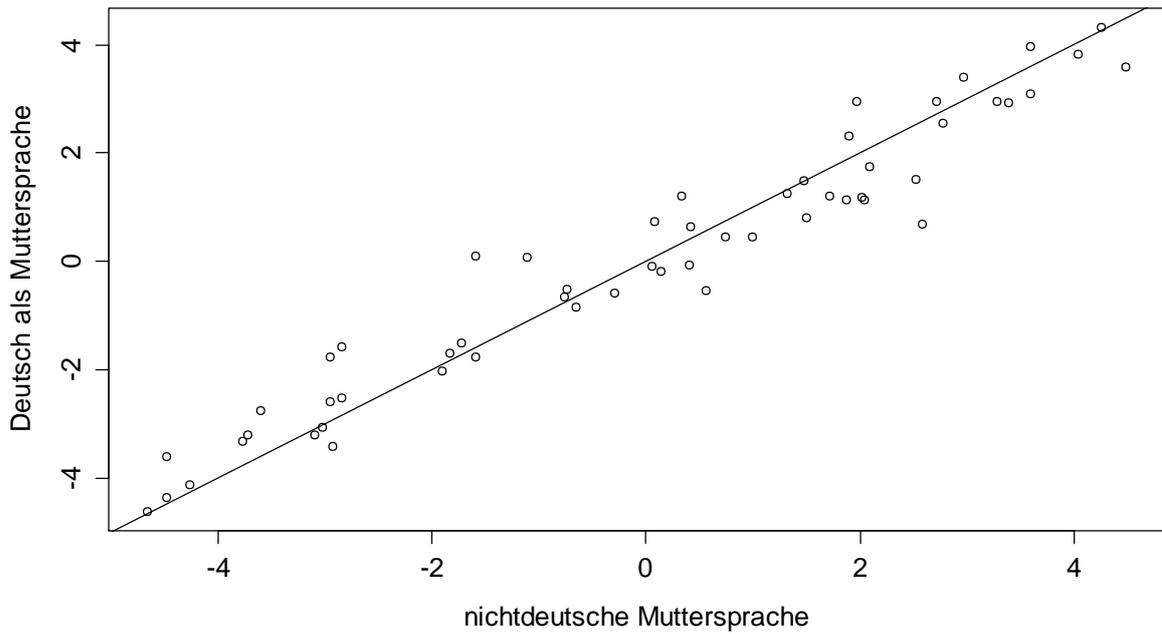
Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	P
Score	147,585	57	83,541	< 0,001
Geschlecht	63,605	57	83,541	0,255
Muttersprache	67,962	54	81,069	0,096



**Abbildung 5.50:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 vor dem Ausschluss von Aufgaben (n=442).

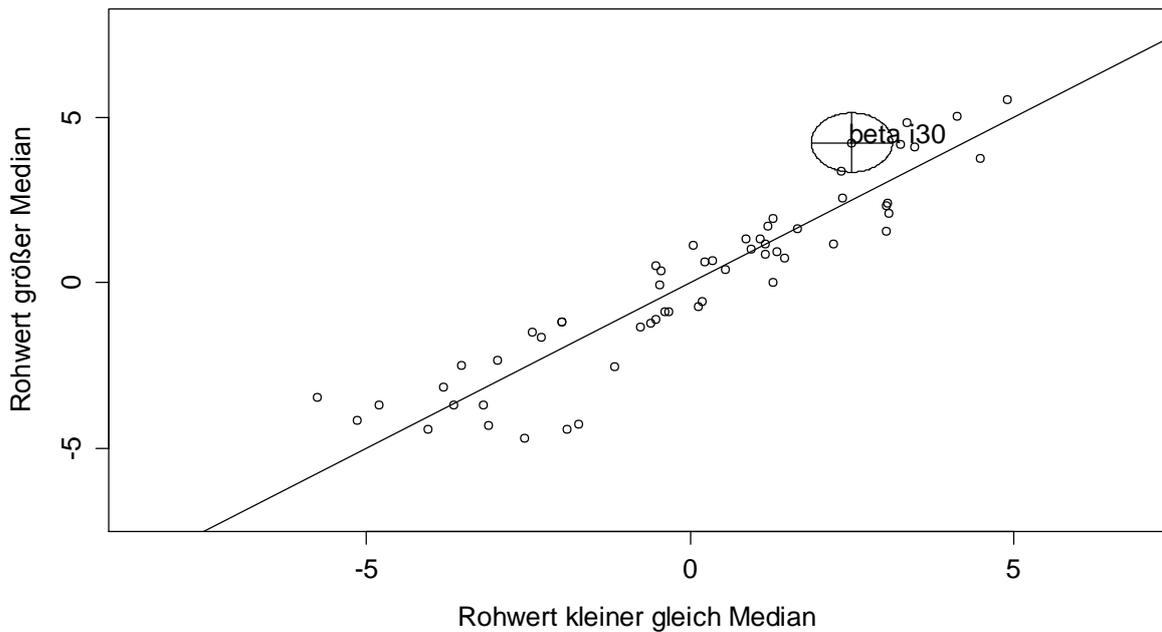


**Abbildung 5.51:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 9 vor dem Ausschluss von Aufgaben (n=442).

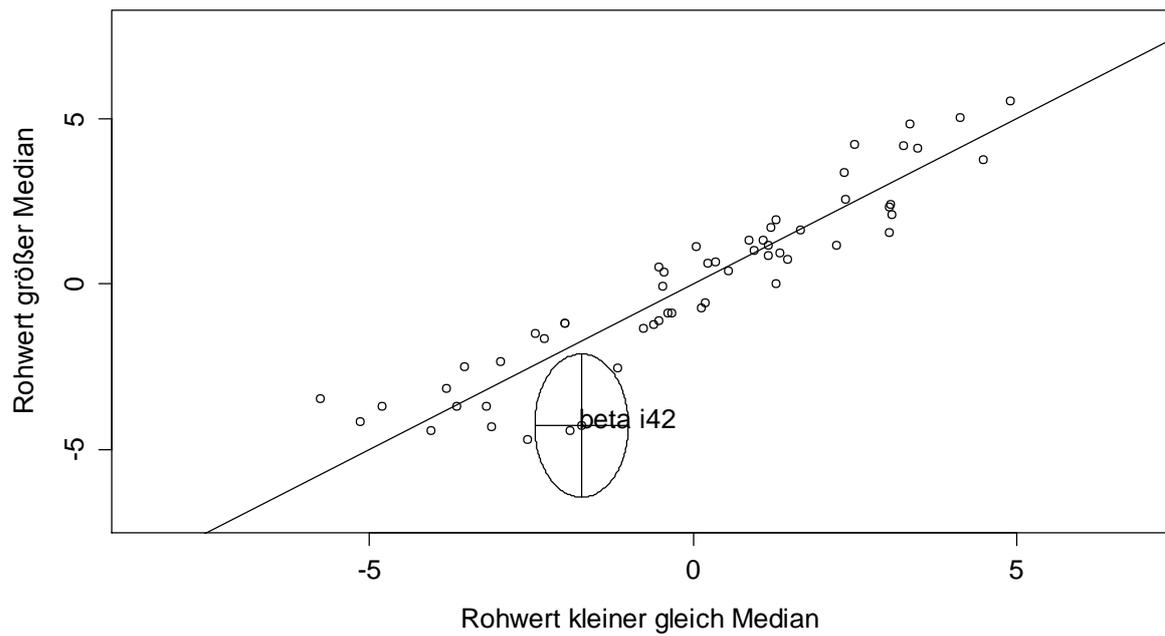


**Abbildung 5.52:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 9 vor dem Ausschluss von Aufgaben (n=442).

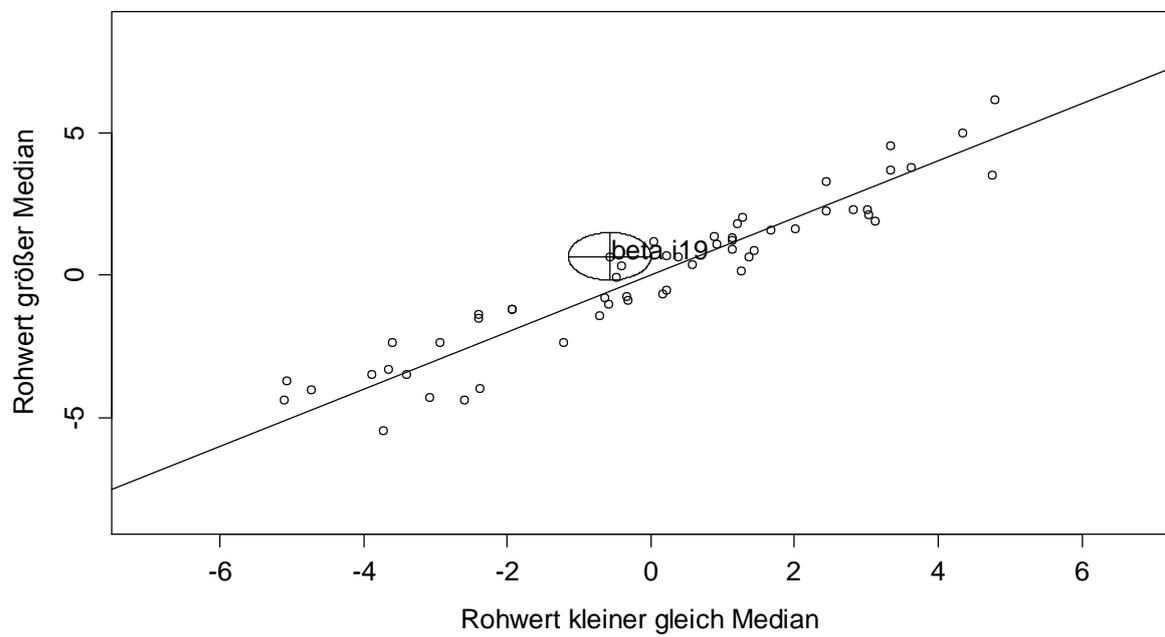
Bei der Graphischen Modellkontrolle für das Teilungskriterium Score waren schrittweise die Items i30, i42, i19, i14 und i10 auffällig. (vgl. Abbildungen 5.53 bis 5.57).



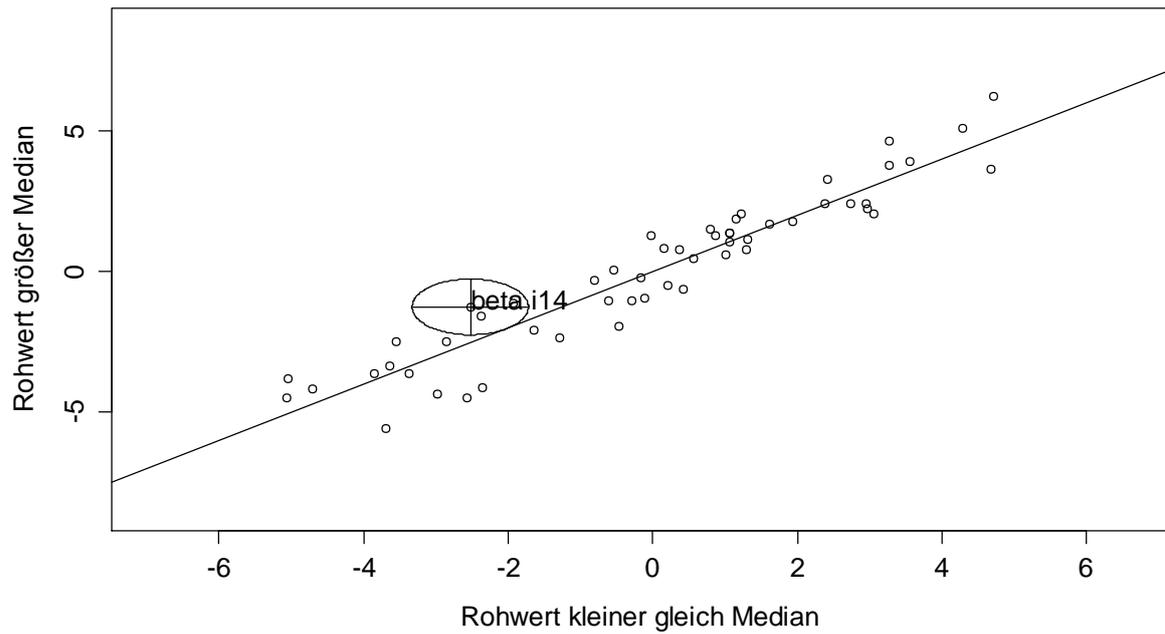
**Abbildung 5.53:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 30 ( $\alpha=0,01$ ); (n=442).



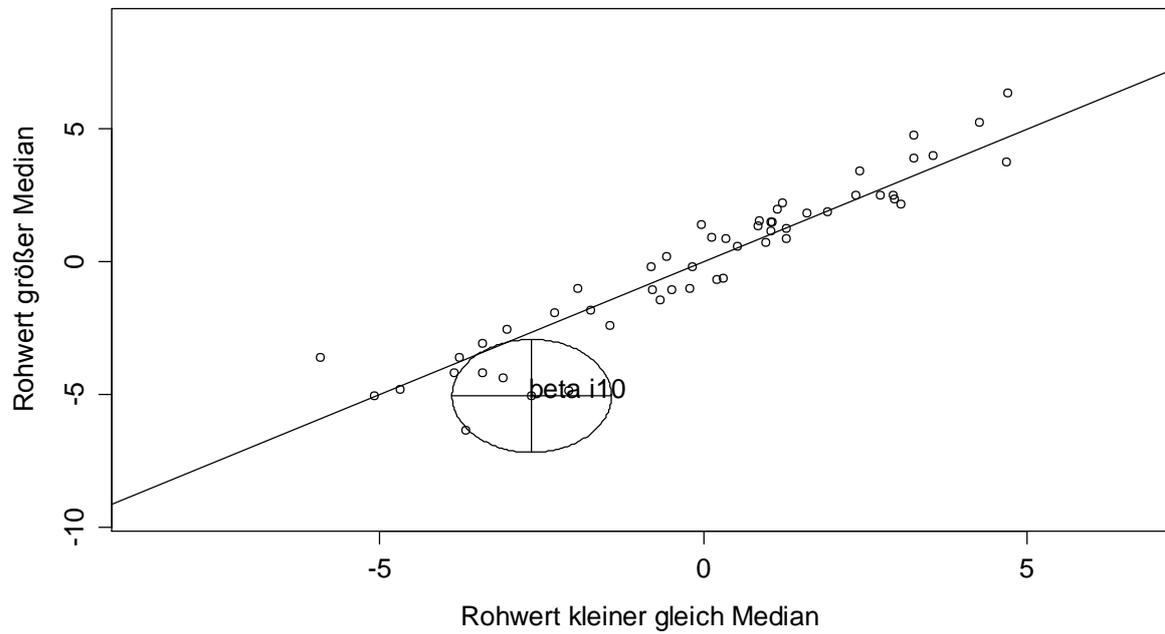
**Abbildung 5.54:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 42 ( $\alpha=0,01$ ); ( $n=442$ ).



**Abbildung 5.55:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 19 ( $\alpha=0,01$ ); ( $n=442$ ).



**Abbildung 5.56:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 14 ( $\alpha=0,01$ ); ( $n=442$ ).



**Abbildung 5.57:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 9 mit Standardschätzfehler des Items 10 ( $\alpha=0,01$ ); ( $n=442$ ).

Nach dem sukzessiven Ausschluss der auffälligen fünf Items (i30, i42, i19, i14 und i10) kann dem UT 9 a-posteriori in der vorgenommenen retrospektiven Analyse für die erhobene Stichprobe Modellgültigkeit zugesprochen werden (siehe Tabelle 5.15).

**Tabelle 5.15:** Darstellung der Ergebnisse der LQTs für die Teilkriterien Score, Geschlecht und Muttersprache des Untertests 9 nach dem Ausschluss von fünf Aufgaben (n=442).

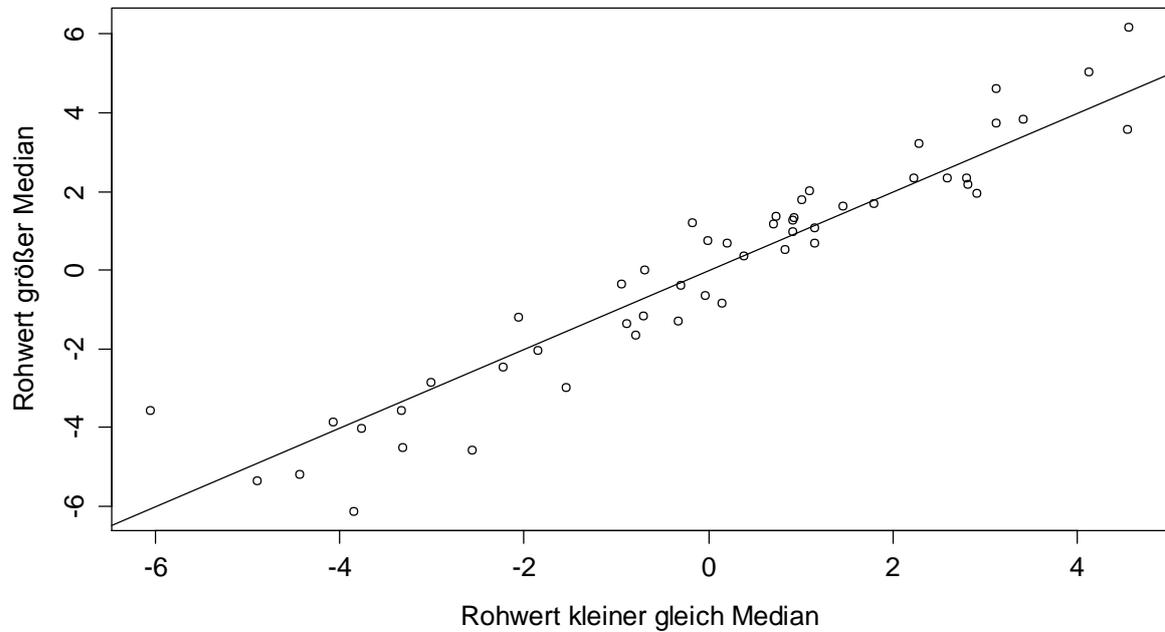
<b>Teilkriterium</b>	$\chi^2$	<b>df</b>	$\chi^2_{\text{krit}} (\alpha=0,01)$	<b>p</b>
Score	66,299	55	82,292	0,069
Geschlecht	58,567	52	78,616	0,247
Muttersprache	43,087	49	74,919	0,711

Keine Aufgabe die für den AID 3 (Version 1.1) aus inhaltlichen Überlegungen ausgeschlossen wurde, ist bei der Reanalyse des AID 2 (Version 2.1 und Version 2.2) teststatistisch auffällig. Bei den Aufgaben die aufgrund der Fragestellung ausgeschlossen werden müssen handelt es sich um i10, i14, i19, i30 und i42 (vgl. Tabelle 5.16).

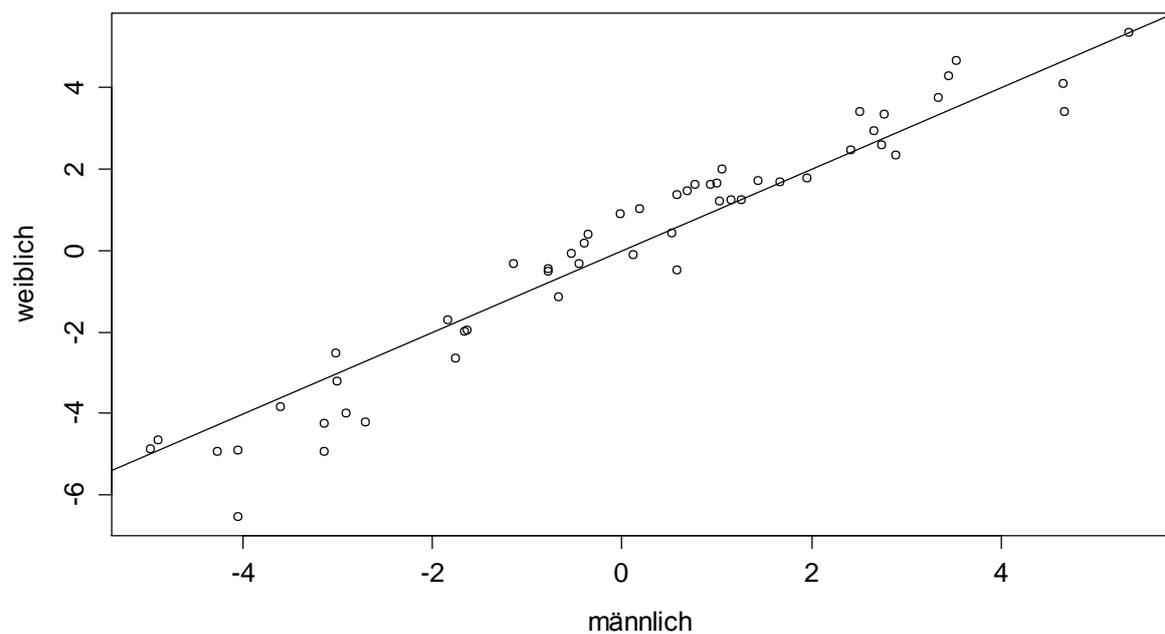
**Tabelle 5.16:** Überblick zu den Items des Untertests 9 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.

<b>Untertest</b>	<b>Items die nicht in den AID3 übernommen wurden</b>	<b>Items die nicht analysiert werden konnten</b>	<b>Items die ausgeschlossen werden mussten</b>
UT 9	i6	i39	i1
	i17	i45	i3
	i18	i50	i4
	i22	i53	i37
	i29	i57	i39
	i31	i39	i44
	i33		i48
	i35		i10
			i14
			i19
			i30
			i42

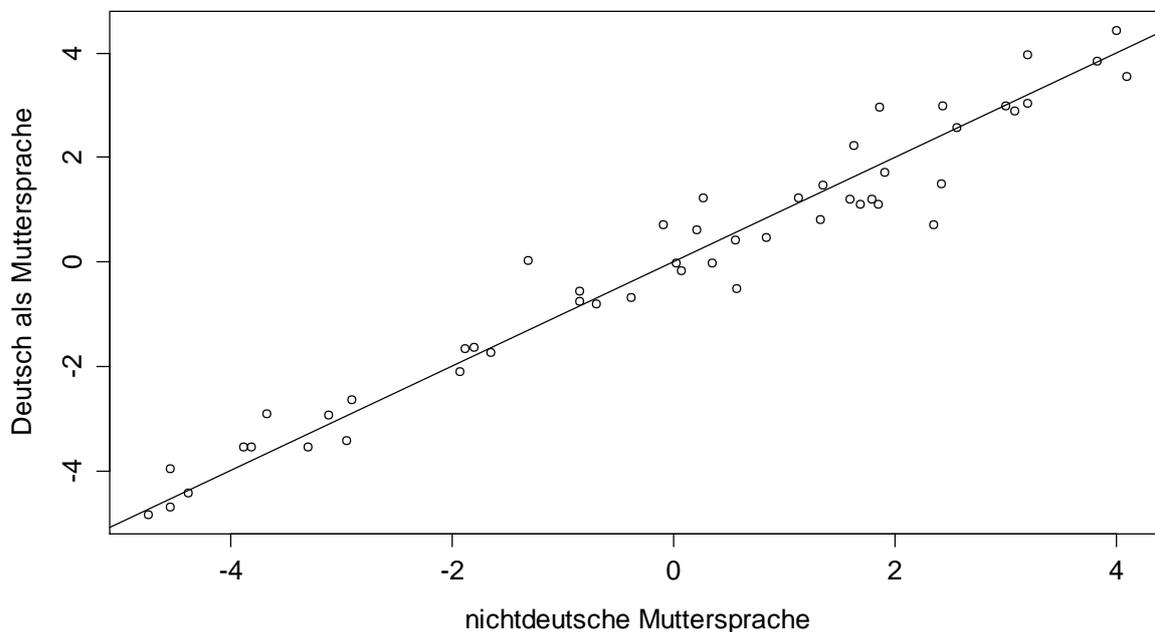
Nachfolgend werden die Grafischen Modellkontrollen der gewählten Teilungskriterien für den a-posteriori gültigen Itempool des Untertests 9 nach dem Ausscheiden von fünf Items dargestellt. (vgl. Abbildungen 5.58 bis 5.60)



**Abbildung 5.58:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 9 nach dem Ausscheiden von fünf Items.



**Abbildung 5.59:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 9 nach dem Ausscheiden von fünf Items.



**Abbildung 5.60:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 9 nach dem Ausscheiden von fünf Items.

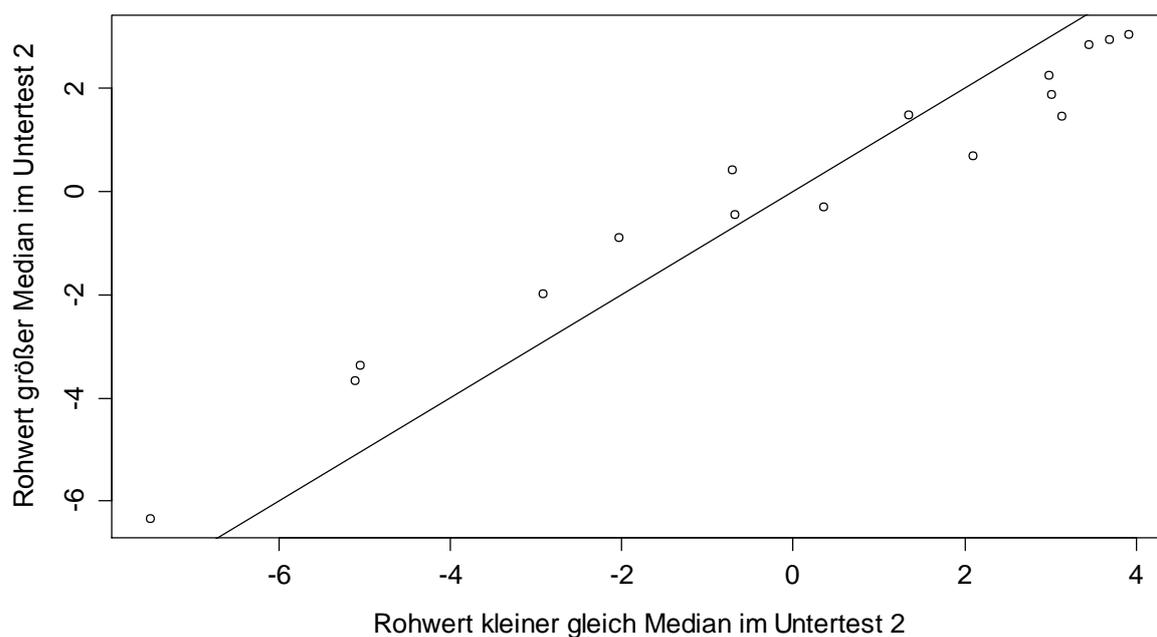
### 5.7 Untertest 10 Analysieren und Synthetisieren-abstrakt

Im Untertest 10 mussten Tpn 0113 und 0224 wegen fehlender Antworten ausgeschlossen werden. Die Aufgaben i8 und i7 konnten so, wie vier sogenannte *warming up* Items nicht analysiert werden. Tabelle 5.14 zeigt das Ergebnis des ersten LQTs sowie die Anzahl schätzbarer Items.

**Tabelle 5.17:** Darstellung des Ergebnisses des LQTs des Untertests 10 für das Teilungskriterium Score im Untertest 2 (n=443).

Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	p
Score	9,173	15	30,579	0,868

Abbildung 5.50 zeigt die Grafische Modellkontrolle für das gewählte Teilungskriterium. Im UT 10 mussten keine Aufgaben ausgeschlossen werden.



**Abbildung 5.61:** Grafische Modellkontrolle des Untertests 10 mittels des Teilungskriteriums Score im UT 2 (n=443).

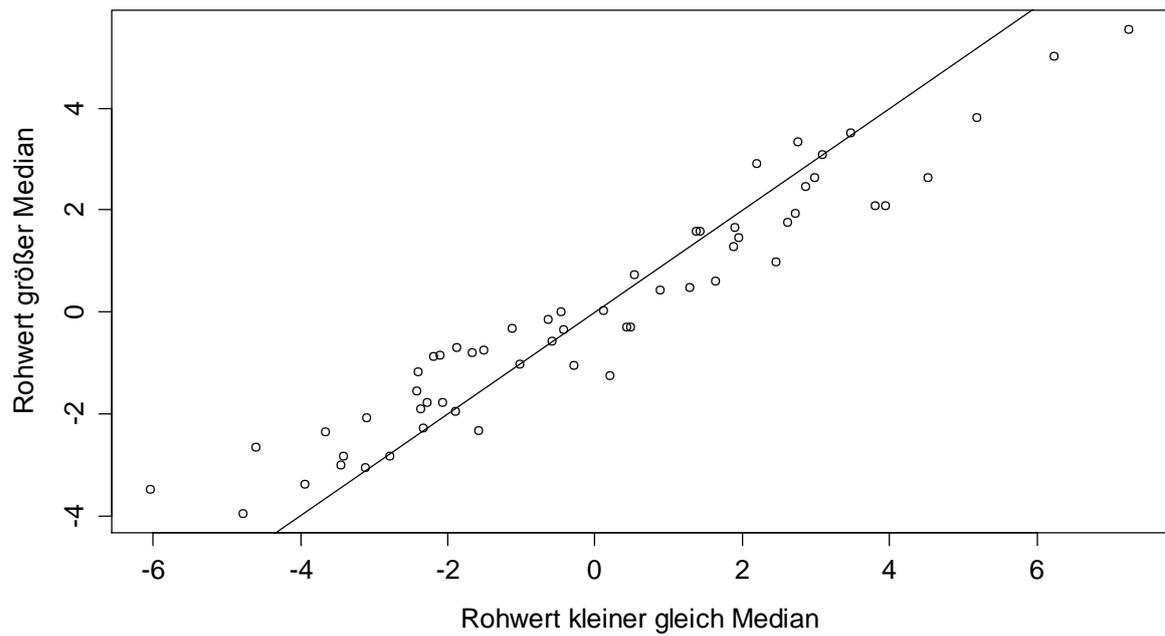
## 5.8 Untertest 11 Soziales Erfassen und Sachliches Reflektieren

Im Untertest 11 mussten die Tpn 0224, 0113, 0247 und 0340 wegen fehlender Antworten ausgeschlossen werden. Für die gewählten Teilungskriterien sind folgende Aufgaben nicht schätzbar: i1, i5, i36, i37, i38, i39 und i40. Tabelle 5.18 zeigt signifikante Ergebnisse der ersten LQTs für das Tk Score und das Tk Muttersprache sowie die Anzahl schätzbarer Items vor dem Ausschluss des ersten Items.

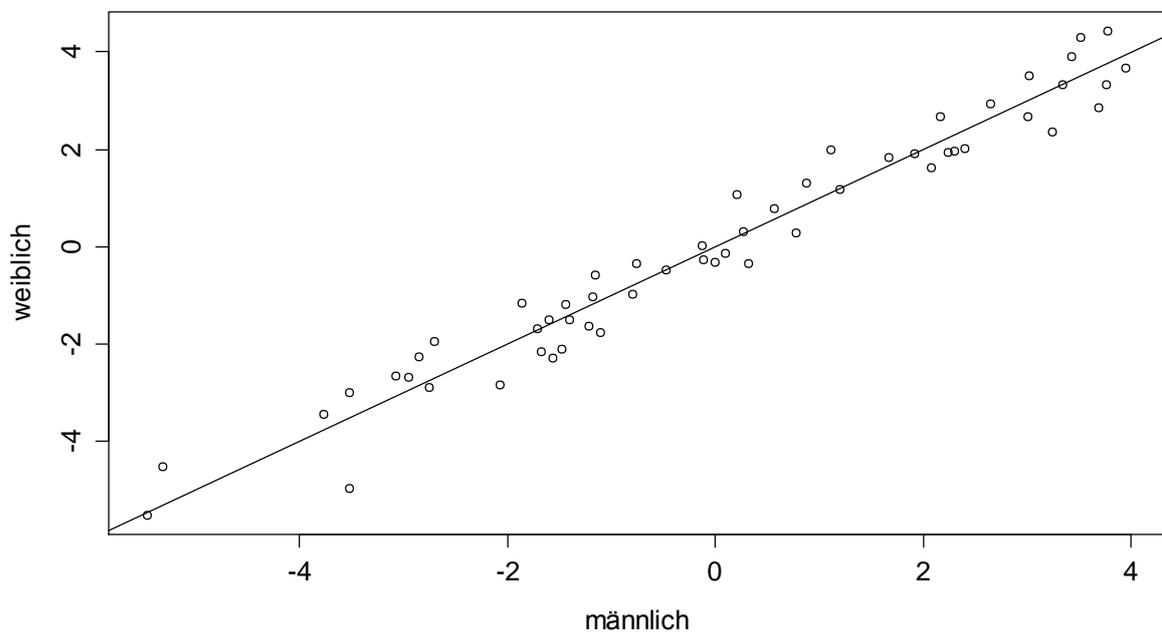
**Tabelle 5.18:** Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 11 vor dem Ausschluss von Aufgaben (n=441).

Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	p
Score	98,949	57	84,732	<0,001
Geschlecht	58,373	55	82,292	0,352
Muttersprache	86,910	54	81,069	0,003

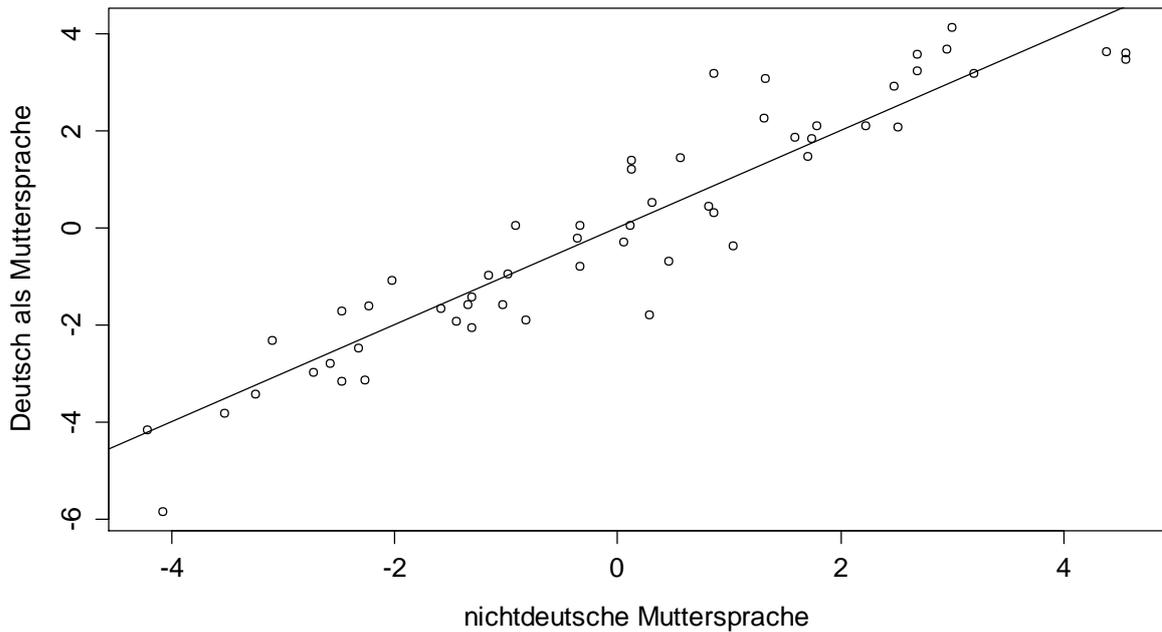
Die Abbildungen 5.62 bis 5.64 zeigen die Grafischen Modellkontrollen für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 11 vor dem Ausschluss von Aufgaben.



**Abbildung 5.62:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 vor dem Ausschluss von Aufgaben (n=441).

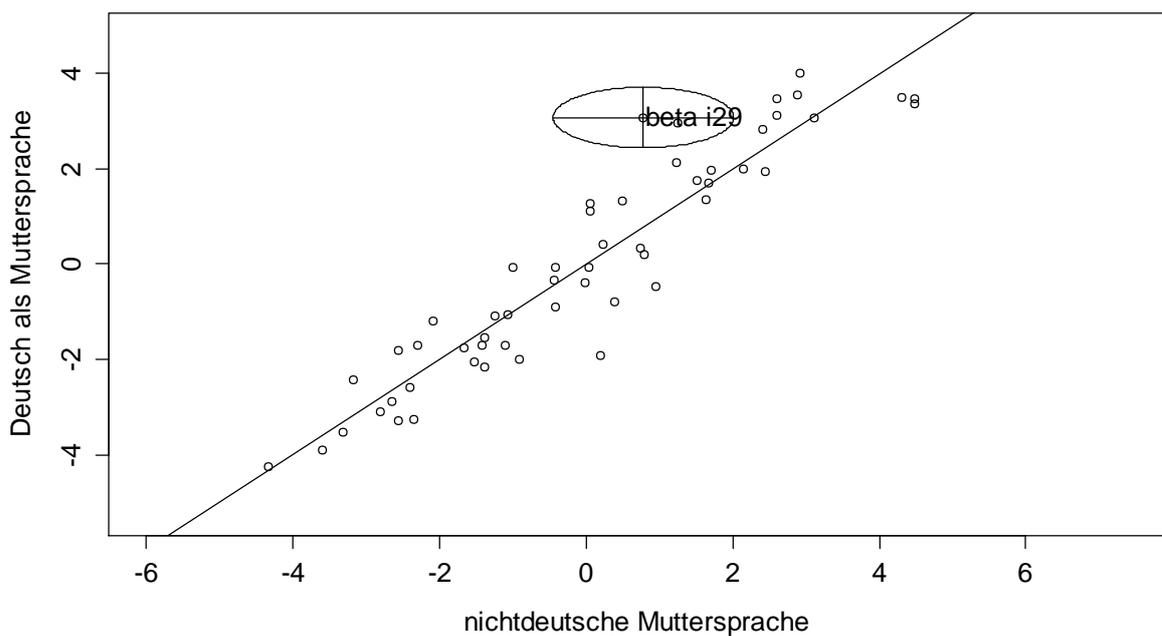


**Abbildung 5.63:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den UT 11 vor dem Ausschluss von Aufgaben (n=441).

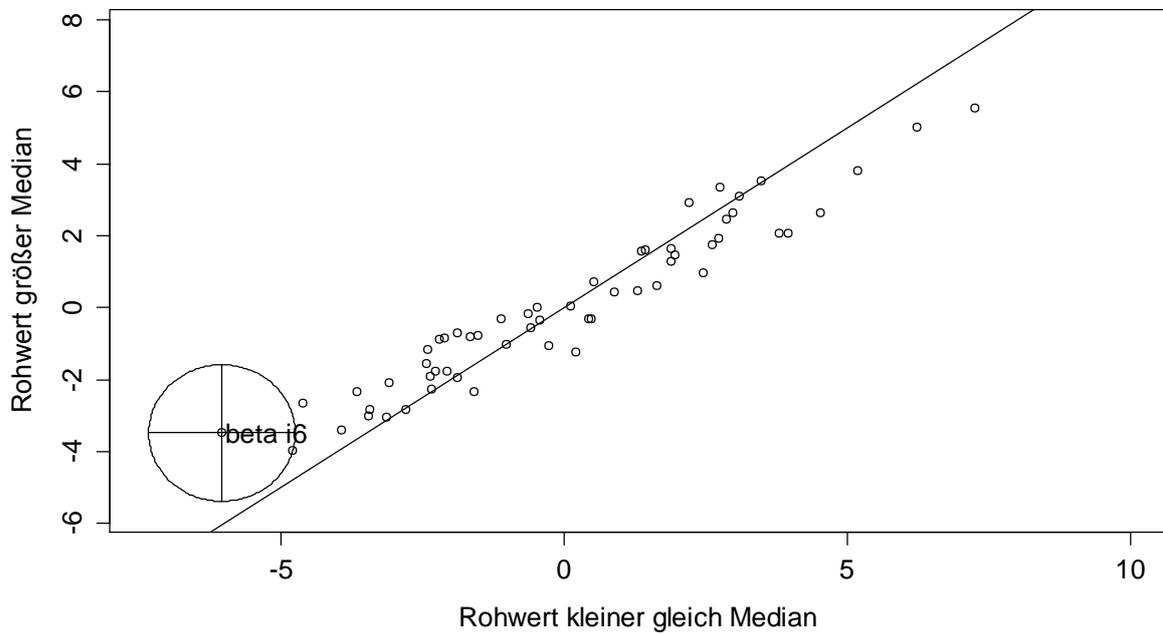


**Abbildung 5.64:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den UT 11 vor dem Ausschluss von Aufgaben (n= 441).

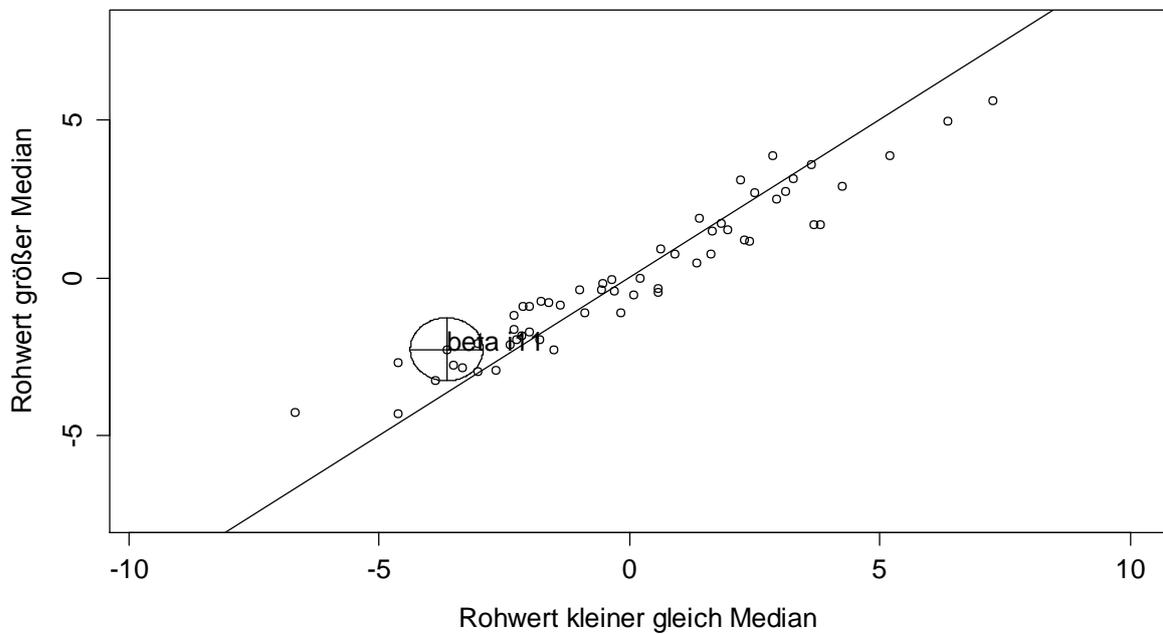
Nach dem signifikanten Ergebnissen des LQTs wurden bei der Grafischen Modellkontrolle für die Teilungskriterien Score die Items i6, i7, i11 und das Linking-Item i46 auffällig. Für das Tk Sprache wurde das Item i29 auffällig (vgl. Abbildungen 5.65 bis 5.69).



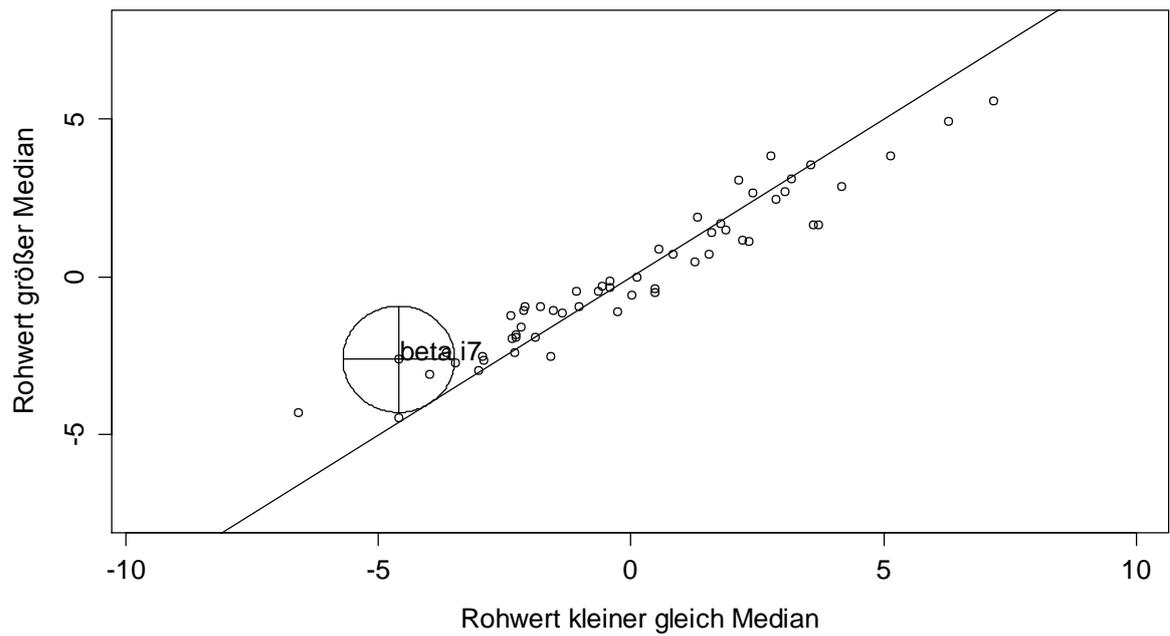
**Abbildung 5.65:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 29 ( $\alpha=0,01$ ); ( $n=441$ ).



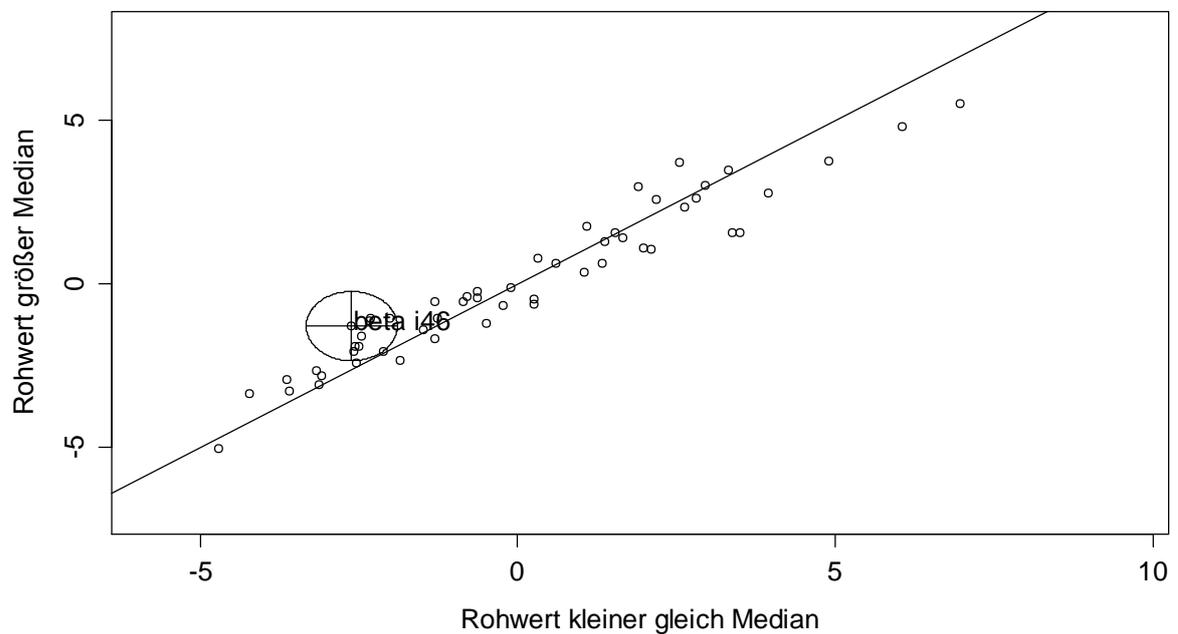
**Abbildung 5.66:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 6 ( $\alpha=0,01$ ); ( $n=441$ ).



**Abbildung 5.67:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 11 ( $\alpha=0,01$ ); ( $n=441$ ).



**Abbildung 5.68:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 7 ( $\alpha=0,01$ ); ( $n=441$ ).



**Abbildung 5.69:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den UT 11 mit Standardschätzfehler des Items 46 ( $\alpha=0,01$ ); ( $n=441$ ).

Nach dem schrittweisen Ausschluss von fünf Items kann dem UT 11 a-posteriori in dieser Reanalyse Modellgültigkeit zugesprochen werden. Die LQTs der drei Teilungskriterien für den UT 11 nach dem Ausschluss von fünf Items zeigt Tabelle 5.19.

**Tabelle 5.19:** Darstellung der Ergebnisse der LQTs für die Teilungskriterien Score, Geschlecht und Muttersprache des Untertests 11 nach dem Ausscheiden von fünf Aufgaben (n=441).

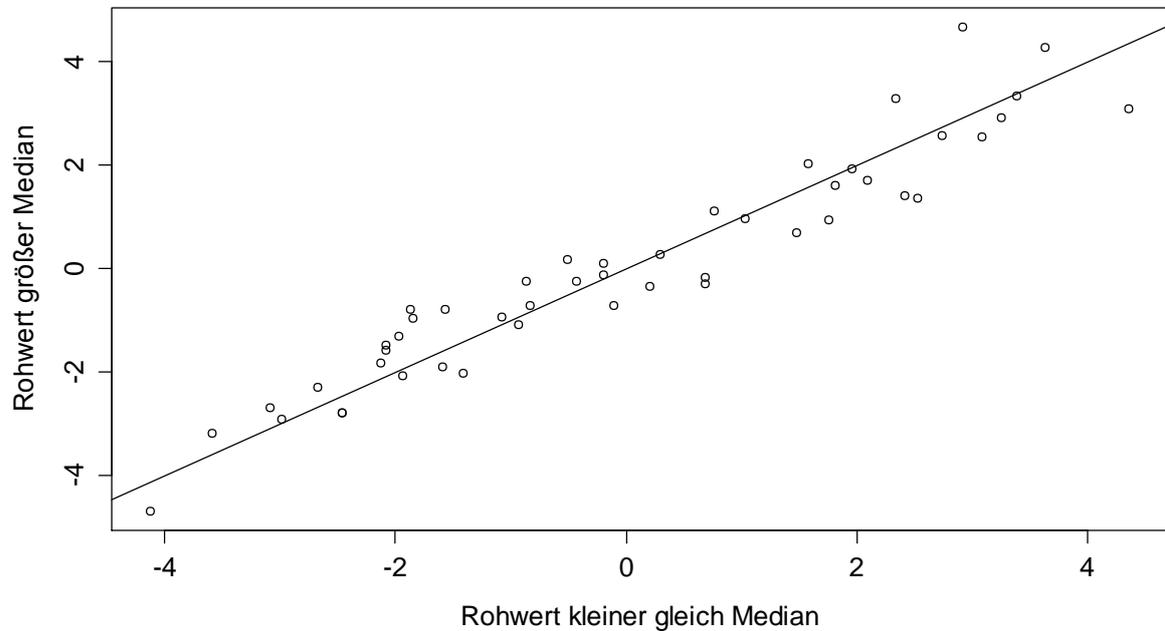
<b>Teilungskriterium</b>	$\chi^2$	<b>df</b>	$\chi^2_{\text{krit}} (\alpha=0,01)$	<b>p</b>
Score	65,112	47	72,443	0,041
Geschlecht	52,872	50	76,154	0,364
Muttersprache	63,587	49	74,920	0,079

Tabelle 5.20 zeigt den im Überblick welche Aufgaben nicht in den AID 3 (Version 3.1) übernommen wurden, welche Items des AID 2 (Version 2.1 und Version 2.2) nicht analysiert werden konnten und welche ausgeschlossen werden mussten.

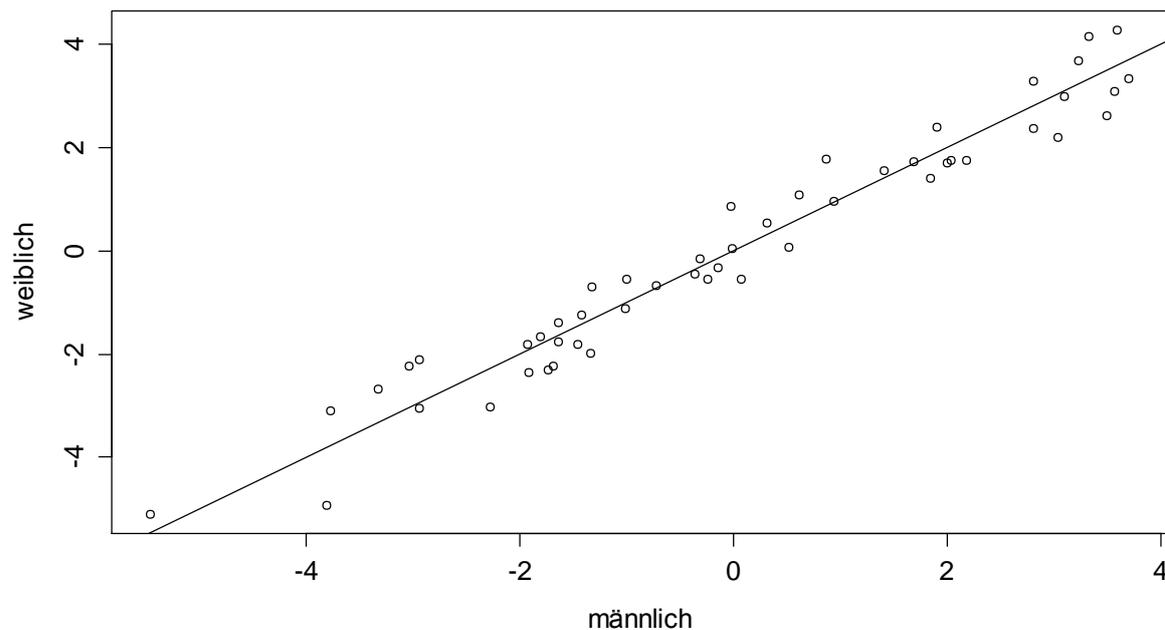
**Tabelle 5.20:** Überblick zu den Items des Untertests 11 im AID 2 (Version 2.1 & Version 2.2) und AID 3 (Version 3.1) nach der Analyse.

<b>Untertest</b>	<b>Items die nicht in den AID3 übernommen wurden</b>			<b>Items die nicht analysiert werden konnten</b>	<b>Items die ausgeschlossen werden mussten</b>
UT 11	i3	i29	i58	i1	i6
	i4	i30	i59	i5	i7
	i12	i35		i36	i11
	i18	i39		i37	i29
	i19	i42		i38	i46
	i21	i43		i39	
	i22	i46		i40	
	i24	i53			
	i28	i55			

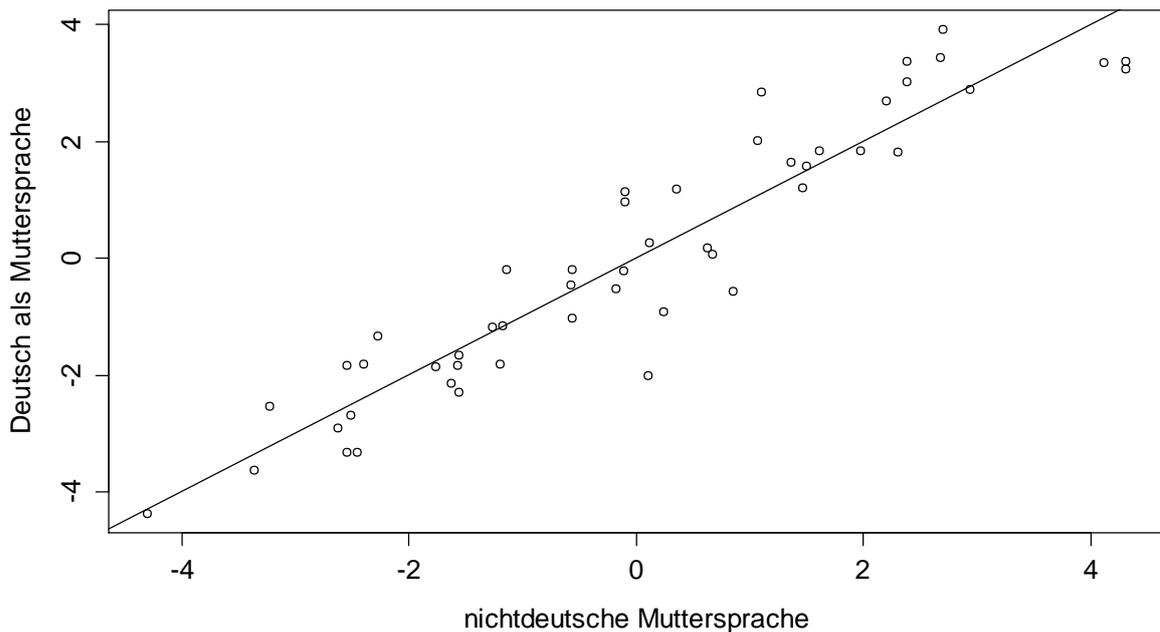
Die Grafischen Modellkontrollen der drei Teilungskriterien für den a-posteriori gültigen Itempool des Untertests 11 können den Abbildungen 5.70 bis 5.72 entnommen werden.



**Abbildung 5.70:** Grafische Modellkontrolle mittels des Teilungskriteriums Score für den a-posteriori gültigen Itempool des Untertests 11 nach dem Ausscheiden von fünf Items (n=441).



**Abbildung 5.71:** Grafische Modellkontrolle mittels des Teilungskriteriums Geschlecht für den a-posteriori gültigen Itempool des Untertests 11 nach dem Ausschluss von fünf Item (n=441).



**Abbildung 5.72:** Grafische Modellkontrolle mittels des Teilungskriteriums Sprache für den a-posteriori gültigen Itempool des Untertests 11 nach dem Ausscheiden von fünf Items (n=441).

## 6. Diskussion

Im Rahmen der Diskussion wird eine Itemanalyse jener Aufgaben, welche bei der Grafischen Modellkontrolle nach *Rasch* auffällig wurden, vorgenommen. Dabei ist jedoch zu beachten, dass laut neuesten Ergebnissen von Kubinger et al. (in Vorb.) alle *Likelihood*-Quotienten-Tests nach *Andersen* nicht signifikant sind und daher kein Item auszuschließen gewesen wäre. Aus Gründen der Testsicherheit konnten die betreffenden Aufgaben nicht widergegeben werden.

Im folgenden Kapitel steht der besseren Lesbarkeit geschuldet die Bezeichnung AID 2 für den AID 2 (Version 2.1) und AID 2 (Version 2.2). Der Begriff AID 3 wird anstelle der vollständigen Bezeichnung AID 3 (Version 3.1) verwendet.

## 6.1 Itemanalyse für den UT 1

Im UT 1 waren elf der fünfundzwanzig nicht analysierbaren Aufgaben *Linking-Items*. Die fehlenden *Linking-Items* können bewirken, dass einige Aufgaben nicht mehr mittels LQT vergleichbar sind, da Teilstichproben von Items entstehen können, die nicht mehr verlinkt sind, oder Teilstichproben von Personen entstehen, die alle oder keine der *Linking-Items* gelöst haben.

Im UT 1 wurden 14 Items schrittweise ausgeschlossen um a-posteriori Modellgültigkeit zu erreichen. Von diesen 14 Aufgaben wurden zwei Aufgaben auch für den AID 3 ausgeschlossen (i54 und i57).

In einem ersten Analyseschritt wurde das Item i51 ausgeschlossen. Im Tk Muttersprache zeigte sich für das Item eine Differenz der Parameterschätzungen in beiden Teilstichproben von 2,693. Es war für die Tpn mit nichtdeutscher Muttersprache schwieriger zu lösen. Möglicherweise liegt dies daran, dass im deutschen Sprachraum der Begriff *WIDIA* alltagssprachlich verwendet wird. *WIDIA* ist eine Abkürzung für "wie Diamant" und ursprünglich ein Markenname, der aber mittlerweile in der Alltagssprache für ganze Werkzeugproduktgruppen eine synonyme Anwendung findet. Es könnte daher sein, dass Tpn mit deutscher Muttersprache eher an die gesuchten alternativen Einsatzmöglichkeiten von Diamanten denken.

Ein anderer Erklärungsansatz bezieht sich auf die Verwendung des Wortes *Schmuck* im Item. Es wird angenommen dass anstelle des Überbegriffes *Schmuck* in der Alltagssprache eher Wörter wie Kette, Ohrring oder Ring verwendet werden. Wenn weiters angenommen wird, dass häufig verwendete Begriffe der Alltagssprache für Tpn mit nichtdeutscher Muttersprache geläufiger sind, würde dies implizieren dass der Begriff *Schmuck* diesen Tpn eher nicht bekannt ist.

Bei Item i48 zeigte sich für das Tk Muttersprache mit einer Parameterdifferenz von 2,382 der gegenteilige Effekt. Dieses Item war für Tpn mit nichtdeutscher Muttersprache leichter zu lösen. Der fremdsprachige Name, welcher im Item vorkommt, könnte für Tpn mit deutscher Muttersprache schwieriger zu lesen sein als für fremdsprachige Tpn, und diese könnten deshalb besser abschneiden. Möglicherweise sind für Tpn mit deutscher Muttersprache Namen deutschsprachiger Künstler einprägsamer, für fremdsprachige Tpn hingegen internationale Namen.

In einem weiteren Analyseschritt war Aufgabe i50 auszuschließen. Im Tk Muttersprache zeigte das Item eine Parameterdifferenz von 2,223. Die Aufgabe ist für Tpn mit deutscher Muttersprache schwieriger zu lösen. Womöglich verlassen sich Kinder und Jugendliche mit deutscher Muttersprache bei Fragen bezüglich der Gesundheit mehr auf ihre Eltern als Kinder und Jugendliche mit nichtdeutscher Muttersprache.

Das Item i20 war mit einer Parameterdifferenz von 1,742 bei einer Range der Parameterschätzungen von 17,007 für leistungsschwächere Tpn einfacher zu lösen als für leistungsstärkere. Möglicherweise wissen leistungsfähigere Tpn prinzipiell mehr über kriegerische Auseinandersetzungen im historischen Kontext. Genau dieses Mehr an Wissen könnte dazu führen, dass leistungsfähigere Tpn Kriege, bei denen mehrere Länder beteiligt waren, fälschlicher Weise als Weltkriege bezeichnen.

Das Item i13 zeigte in den Teilungskriterien Score (Parameterdifferenz von -1,382) und Sprache (Parameterdifferenz von 1,360) eine teststatistische Auffälligkeit. Dieser war aber keine praktische Relevanz beizumessen, da die Parameterdifferenz im Tk Score kleiner als ein Zehntel der Range der Parameterschätzungen war. Für das Tk Score beträgt die Spannweite der Parameterschätzungen 17,467, für das Tk Sprache 13,185.

Die Aufgabe i42 zeigte für das Tk Score eine Parameterdifferenz von 4,499 und war für leistungsschwächere Tpn schwieriger zu beantworten als für leistungsstärkere. Als Ursache für diesen Unterschied könnte angenommen werden, dass es leistungsfähigeren Tpn eher gelingt, einen Rückschluss auf das Ausgangsprodukt herzustellen, wodurch nicht nur Sachkenntnisse erfasst würden.

Die Aufgabe i55 war mit einer Parameterdifferenz von 1,584 bei einer Range von 10,924 der Parameterschätzungen für das Tk Sprache auffällig. Es war für Tpn mit deutscher Muttersprache leichter zu lösen als für Tpn mit nichtdeutscher Muttersprache. Bei diesem Item rückten vor allem die korrekten Antwortmöglichkeiten in das Zentrum der Itemanalyse. Die Antworten bestehen aus komplexen Wortzusammensetzungen, bei denen Tpn mit nichtdeutscher Muttersprache benachteiligt sein könnten.

Das Item i12 war mit einer Parameterdifferenz im Tk Geschlecht von 1,423 bei einer Range aller Parameter von 12,815 für männliche Tpn leichter. Dieser Unterschied könnte Geschlechterstereotyp begründet werden. Burschen interessieren sich demnach mehr für die Eisenbahn als Mädchen.

Die Aufgabe i16 zeigte mit einer Parameterdifferenz von 1,472 in Bezug zu der Spannweite der Parameterschätzungen von 11,361 für das Tk Sprache, dass die Frage in der Stichprobe für Kinder mit Deutsch als Muttersprache schwieriger zu beantworten war. Möglicherweise verbinden Tpn mit deutscher Muttersprache Weihnachten mit der kalten Jahreszeit und weniger mit einem bestimmten Monat. Für Tpn mit nichtdeutscher Muttersprache könnte das Weihnachtsfest jedoch ein Festtag sein, den sie aus ihrem Kulturkreis nicht kennen, und da es ein zusätzliches Fest in ihrem Alltag ist, könnte es besser eingeprägt werden.

Bei Item i26 zeigte die Parameterdifferenz von 1,457 bei einer Range der Parameter von 12,879 ein für Burschen leichter zu lösendes Item. Dem Geschlechterstereotyp entsprechend könnten weibliche Tpn mit dem Thema Steuern seltener konfrontiert werden, weshalb die Aufgabe für sie schwieriger zu lösen sein könnte.

Im vorletzten Analyseschritt wurde die Aufgabe i23 auffällig. Das Item war mit einer Parameterdifferenz von 1,766 bei einer Range der Parameterschätzungen von 14,462 im Tk Score für leistungsschwächere Kinder und Jugendliche leichter als für leistungsstärkere. Leistungsfähigere Kinder könnten dazu tendieren zu glauben, dass sie die Aufgabe beantworten können, weil sie ohnehin häufig mit den gesuchten Namen konfrontiert werden. Nichtdeutschsprachigen Tpn sind die gesuchten Namen hingegen nicht so geläufig, weshalb sie eher dazu gezwungen sein könnten, diese auswendig zu lernen. Dieses Wissen wäre demnach besser gespeichert. Gleichzeitig läge der Gedanke nahe, dass nichtdeutschsprachige Tpn aufgrund der Immigration mit dem Thema der Nachbarländer intensiver konfrontiert sind und dieses Wissen eine bedeutendere Rolle spielen könnte.

Zuletzt wurde das Item i31 ausgeschlossen. Die Berechnungen ergeben eine Parameterdifferenz von -3,054 für das Tk Score. Das Item ist für leistungsstärkere Tpn leichter zu beantworten als für leistungsschwächere. Dieses Item wurde allerdings nur von 45 Tpn bearbeitet und von 9 Tpn gelöst. Der Unterschied in der Schwierigkeit des Items könnte darin liegen, dass sich leistungsfähigere Tpn auch für die Hintergründe, das Warum, interessieren, wohingegen leistungsschwächere Tpn dieses Verhalten nicht so häufig zeigen könnten.

## 6.2 Itemanalyse für den UT 2

Auf Grund der Analysen war im UT 2 die Aufgabe i8 als auffällig zu bewerten. Die Differenz der Itemleichtigkeitsparameter von Item i8 im betrug 0,414, einer Range der Parameterschätzungen für das Tk Score 5,930. Der Ausschluss des Items wäre aufgrund des Effekts, der kleiner als ein Zehntel aller Parameterschätzungen ist, in der Praxis nicht rechtfertigbar. Ein so geringer Effekt wäre inhaltlich nicht relevant (vgl. Kapitel 4 und Kubinger, 2005). Das Item sollte im Itempool verbleiben.

## 6.3 Itemanalyse für den UT 3

Im UT 3 waren die Items i23 und i24 auffällig. Das Item i24 wurde aus inhaltlichen Gründen auch für den AID 3 ausgeschlossen. Die Ergebnisse der teststatistischen Analysen zeigten, dass das Item i23 mit einer Parameterdifferenz von 1,889 und einer Range aller Parameterschätzungen im Tk Score von 17,035 sowie im Tk Sprache mit einer Parameterdifferenz von 1,769 bei einer Range der Parameter von 15,763 ebenfalls auszuschließen war. Das Item i23 war für leistungsschwächere Testpersonen und Personen mit nichtdeutscher Muttersprache leichter zu lösen. Die Aufgabe ist sprachlich einfach gestaltet. Leistungsstärkere Tpn könnten deshalb zu Flüchtighkeitsfehlern tendieren. Möglicherweise verkomplizieren sie die Fragestellung aber auch, indem sie an komplexe Textaufgaben im Unterricht denken.

Womöglich interessieren sich leistungsschwächere Tpn eher für Freizeitaktivitäten und den Zirkus. Sie könnten die Aufgabe deshalb aufmerksamer bearbeiten.

## 6.4 Itemanalyse für den UT 4

Bei der Grafischen Modellkontrolle im UT 4 war für das gewählte Teilungskriterium (Score im UT 2) keines der analysierbaren Items auffällig. Es ist jedoch anzumerken, dass von den achtzehn Aufgaben in der erhobenen Stichprobe sieben nicht analysiert werden konnten (vgl. Abbildung 5.36).

## 6.5 Itemanalyse für den UT 6

Von den im UT 6 in der vorliegenden Analyse ausgeschlossenen Items ist ein Item (i23) mit jenen Aufgaben ident, welche im AID 3 aus inhaltlichen Gründen ausgeschlossen wurden. Das Item i23 zeigt eine Parameterdifferenz von 1,256 bei einer Range der Parameterschätzungen von 9,239 im Tk Score und ist für die Teilstichprobe der leistungsschwächeren Tpn leichter zu lösen. Für dieses Ergebnis konnte keine plausible Erklärung gefunden werden.

## 6.6 Itemanalyse für den UT 9

Für den UT 9 ist generell festzuhalten, dass keine Aufgabe die für den AID 3 aus inhaltlichen Überlegungen ausgeschlossen wurde, auch bei der retrospektiven Analyse des AID 2 teststatistisch auffällig war.

Die Aufgabe i30 war mit einer Parameterdifferenz von 1,743 und einer Range der Parameterschätzungen von 11,294 im Tk Score für die Teilstichprobe der leistungsschwächeren Tpn leichter zu lösen. Leistungsschwächere Tpn könnte eine weniger komplexe Denkweise dabei unterstützen eine Lösung zu erreichen.

Das Item i42 war mit einer Parameterdifferenz von -2,495 für die leistungstärkeren Kinder leichter zu lösen. Eine Begründung könnte darin liegen, dass leistungstärkere Kinder mehr am Ablauf des Arbeitsalltages der Eltern teilhaben als leistungsschwächere Kinder. Dadurch wären Arbeitstage und Urlaubstage in deren Bedeutungen stärker voneinander getrennt, und eine Lösung der Aufgabe könnte leistungsfähigeren Kindern und Jugendlichen dadurch leichter gelingen.

Das Item i19 war mit einer Parameterdifferenz von 1,229 und einer Range der Parameter von 11,588 im Tk Score für die leistungsschwächeren Tpn einfacher zu lösen. Auch bei dieser Aufgabe könnte die Erklärung für die Unterschiede zwischen den Teilstichproben wie bei Aufgabe i30 darin zu suchen sein, dass leistungsschwächere Tpn eine weniger komplexe Denkweise an den Tag legen und deshalb zu einer einfachen Lösung gelangen könnten.

Das Item i14 war mit einer Parameterdifferenz von 1,259 bei einer Range der Parameterschätzungen von 11,834 im Tk Score für die leistungsschwächeren Tpn einfacher zu lösen. Für die korrekte Antwort muss an das Lesen gedacht werden. Leistungstärkere Tpn denken möglicherweise eher zu komplex und könnten deshalb weniger häufig zu einer korrekten Lösung der Aufgabe gelangen.

Das Item i10 war mit einer Parameterdifferenz von -2,409 im Tk Score für die leistungstärkeren Tpn leichter zu lösen. Der Begriff *Laterne* könnte in der Alltagssprache nicht mehr so häufig verwendet werden, so dass er Kindern und Jugendlichen nicht mehr bekannt ist. Es wäre möglich, dass leistungstärkere Kinder den Begriff aus der Literatur kennen und deshalb bei der Beantwortung der Frage im Vorteil sind.

#### 6.7 Itemanalyse für den UT 10

Bei der Grafischen Modellkontrolle im UT 10 waren für das gewählte Teilungskriterium (Score im UT 2) keine Items auffällig.

#### 6.8 Itemanalyse für den UT 11

Für den UT 11 zeigte die Betrachtung der Ergebnisse, dass zwei Items (i29 mit einer Parameterdifferenz von 2,296 im Tk Sprache und i46 mit einer Parameterdifferenz 1,321 bei einer Range aller Parameterschätzungen im Tk Score 11,084) sowohl in der retrospektiven Modellprüfung aus teststatistischen Gründen als auch im AID 3 aus inhaltlichen Gründen ausgeschlossen wurden. Zu der Aufgabe i46 ist anzumerken, dass sie an der Spitze in einem Verband von sogenannten getürmten Items (i4, i12, i13 und i16) liegt. Getürmt bedeutet, dass diese Items im Tk Score in der Teilstichprobe der leistungsschwächeren Tpn einen nahezu identen Parameterschätzwert erhalten, jedoch für die Tpn der leistungsfähigeren Teilstichprobe leichter werden.

Von den weiteren Items, die aufgrund der Analysen ausgeschlossen wurden (i6, i11 und i7), war die Aufgabe i6 zwar mit einer Parameterdifferenz von 1,325 für das Tk Score teststatistisch auffällig, aber dieses Ergebnis hat keine praktische Relevanz, da die Differenz nicht größer als ein Zehntel der Range aller Parameterschätzungen (13,292) ist. Selbiges zeigte sich für die Aufgabe i11 mit einer Parameterdifferenz von 1,388 im Tk Score bei einer Range der Parameter von 13,940.

Das Item i7 zeigte eine Parameterdifferenz von 1,988 im Tk Score auf und ist für leistungsschwächere Tpn leichter zu lösen als für leistungsstärkere. Die Ursache für den Unterschied in der Schwierigkeit könnte darin gesehen werden, dass leistungsschwächere Tpn manuellen Tätigkeiten möglicherweise eine höhere Affinität und mehr Interesse entgegenbringen. Durch Beobachtungslernen könnte diesen Tpn bewusster sein, dass bei speziellen Situationen und in einigen Berufen Sicherheitswesten getragen werden.

Abschließend kann festgehalten werden, dass für die neueste Generation des AID (AID 3, Version 3.1) für den UT 1 insgesamt 24 Aufgaben im Vergleich zu den Vorgängerversionen des AID 2 (Version 2.1 und Version 2.2) nicht übernommen wurden. Bei der retrospektiven Analyse wurden 14 Aufgaben auffällig. Zwei der betroffenen Aufgaben wurden dabei sowohl im AID 3 ausgeschlossen als auch in der retrospektiven Analyse des AID 2.

Betreffend den Untertest 2 wurden für den AID 3 drei Aufgaben nicht übernommen. Bei der retrospektiven Analyse des AID 2 wurde eine Aufgabe ausgeschlossen. Diese befindet sich auch unter den nicht in den AID 3 übernommenen Items.

Im UT 3 wurden für den AID 3 zwölf Aufgaben ausgeschlossen. Die Analyse im Rahmen der Fragestellung der Diplomarbeit zeigte, dass zwei Items auffällig wurden, wobei ein Item ident mit den ausgeschlossenen Items des AID 3 war.

Im Untertest 4 wurden sowohl für den AID 3 als auch in der retrospektiven Analyse des AID 2 keine Items ausgeschlossen.

Jener UT, bei dem die meisten Aufgaben für den AID 3 ausgeschlossen wurden, ist der UT 6. Insgesamt fanden 35 Aufgaben keinen Eingang in den AID 3. Bei der Analyse des AID 2 waren hingegen nur sieben Items auszuschließen. Von diesen sieben Aufgaben decken sich sechs Items mit jenen, die im AID 3 ausgeschlossen wurden.

Für den UT 9 wurden im AID 3 13 Items ausgeschlossen, bei der retrospektiven Analyse des AID 2 waren fünf Aufgaben auszuschließen. In diesem Untertest deckten sich als einziger keine der ausgeschlossenen Aufgaben mit dem AID 3.

Für den UT 10 gilt wie für den UT 4, dass keine Aufgaben ausgeschlossen wurden. Im AID 3 wurden nur die vier sogenannten *warming up* Items nicht übernommen, welche bei der retrospektiven Analyse aus teststatistischen Gründen von einer Modellprüfung ausgenommen werden mussten.

Im UT 11 wurden für den AID 3 20 Items ausgeschlossen, bei der retrospektiven Analyse des AID 2 waren fünf Aufgaben auszuschließen, wobei zwei der auszuschließenden Items deckungsgleich waren.

Einschränkend muss festgehalten werden, dass Modellprüfungen in Zukunft nur mehr anhand der neuesten Ergebnisse von Kubinger et al. (in Vorb.) durchzuführen sein werden, da die gewählte Methode der Prüfung der Modellgültigkeit des dichotomen logistischen Modells nach *Rasch* für die gegebene Fragestellung zu signifikanten Ergebnissen führt. Nach Kubinger et al. (in Vorb.) werden nämlich keine LQTs signifikant, weshalb auch keine Aufgaben auszuschließen sind.

## 6.9 Synopsis

Im Fall der untersuchten Stichprobe waren vergleichsweise mehr Items nicht analysierbar als ausgeschlossen werden mussten. Dadurch könnte der Schluss gezogen werden, dass das *branched-testing* hohe Anforderungen an die Modellprüfung stellt. Bei dem Vergleich, der aus teststatistischen Gründen nach der gewählten Analysemethode auszuschließenden Aufgaben im AID 2, mit den aus inhaltlichen Gründen ausgeschlossenen Items für den AID 3, wurden im Folgenden auch jene Aufgaben mit einbezogen, welche im Lauf der retrospektiven testtheoretischen Analyse nicht analysierbar waren. Für diese Items gilt, dass anhand der vorliegenden Daten nicht analysiert werden konnte, ob sie dem Modell von Rasch entsprechen. Sie müssten anhand einer neu gewonnenen Stichprobe untersucht werden. Im Rahmen der vorgenommenen retrospektiven teststatistischen Analyse war dies nicht möglich. Deshalb wurden die nicht analysierbaren Aufgaben gemeinsam mit jenen Aufgaben, die auf Basis der retrospektiven Analyse ausgeschlossen werden mussten, dem Itempool gegenübergestellt, der für den AID 3 auszuschließen war. Dabei wird für den AID 2 nur von auffälligen Items gesprochen und nicht mehr von Items, die auszuschließen waren.

Es ergab sich für den UT 1 der Befund, dass teststatistisch mehr Aufgaben im AID 2 auffällig waren als für den AID 3 ausgeschlossen wurden: für den AID 3 wurden 24 Aufgaben ausgeschlossen. In der retrospektiven Analyse des AID 2 waren 38 Items auffällig, wovon 14 Aufgaben ident mit jenen waren, welche im AID 3 ausgeschlossen wurden.

Im UT 2 wurden für den AID 3 drei Aufgaben ausgeschlossen. In der Reanalyse waren insgesamt zwei Aufgaben auffällig, welche beide deckungsgleich mit den ausgeschlossenen Items des AID 3 waren.

Für den UT 3 zeigt sich, dass bei der teststatistischen Analyse des AID 2 mehr Items auffällig waren, als für den AID 3 ausgeschlossen wurden: teststatistisch waren für den AID 2 insgesamt 14 Items auffällig. Dabei waren sechs der Aufgaben ident mit jenen 12 Aufgaben, welche für den AID 3 ausgeschlossen wurden.

Für den AID 3 wurden im UT 6 insgesamt 35 Items ausgeschlossen. Davon sind 16 ident mit jenen 22 Aufgaben, welche bei der teststatistischen Analyse für den AID 2 auffällig wurden. Hier wurden, wie im UT 2 von inhaltlichen Überlegungen ausgehend mehr Items für den AID 3 ausgeschlossen als in der Reanalyse des AID 2 teststatistisch auffällig wurden.

Dieses restriktive Vorgehen im Zuge des inhaltlich begründeten Iteamausschlusses zeigte sich auch im UT 9. Hier wurden für den AID 3 13 Aufgaben ausgeschlossen, bei der teststatistischen Analyse des AID 2 waren 12 Items auffällig. Interessanter Weise war davon nur ein Item deckungsgleich.

Schließlich bestätigt sich dieser Trend auch im UT 11: für den AID 3 wurden 20 Aufgaben ausgeschlossen, bei der Reanalyse des AID 2 wurden 12 Items auffällig, wobei davon drei Aufgaben ident mit jenen waren, die für den AID 3 ausgeschlossen wurden.

Zusammenfassend konnte daraus geschlossen werden, dass die teststatistische Analyse des AID 2 nur in zwei Untertests (von sechs miteinander verglichenen) zu mehr auffälligen Items führte, als dies für die inhaltlich Begründete Vorgehensweise für den Ausschluss von Aufgaben für den AID 3 der Fall war. Der inhaltlich Begründete Ausschluss von Aufgaben ist in diesem Fall somit strenger bezüglich des Ausschlusses von Aufgaben, als eine teststatistische Analyse.

Dieses Ergebnis deckt sich mit dem Hauptbefund dieser Diplomarbeit: ein teststatistisch begründeter Ausschluss von 34 Aufgaben für den AID 2, im Vergleich zu einem Ausschluss von 107 Aufgaben auf Basis inhaltlicher Überlegungen für den AID 3, für die im Rahmen dieser Arbeit miteinander verglichenen Untertests.

Somit lässt sich belegen, dass bei der Neugestaltung des Adaptiven Intelligenzdiagnostikums für den AID 3, anhand des inhaltlichen Ausschlusses von Aufgaben, sehr strenge Maßstäbe an die Items hinsichtlich der Modellgeltung gelegt werden. Der Vergleich der teststatistisch ausgeschlossenen Aufgaben mit den aus inhaltlichen Gründen ausgeschlossenen Items zeigt, dass die beiden Methoden zu unterschiedlichen, aber sich ergänzenden Ergebnissen führen. Dies unterstreicht den bedeutsamen Aspekt, dass bei der Testkonstruktion nicht nur teststatistische Analysen heranzuziehen sind sondern auch eine Auseinandersetzung mit den inhaltlichen Gesichtspunkten notwendig ist. Dem AID 3 ist eine strenge Itemselektion zugunsten einer entsprechenden Modellgültigkeit zu attestieren, obwohl dadurch auf weit mehr Items verzichtet werden musste als teststatistisch notwendig. Und dies trotz des Umstandes, dass, wie bereits an mehreren Stellen in dieser Arbeit erwähnt, nach neuesten Erkenntnissen von Kubinger et al. (in Vorb.), die in dieser Diplomarbeit gewählte methodische Vorgehensweise fälschlicher Weise dazu, dass überhaupt Items ausgeschlossen wurden, da nach Kubinger et al. (in Vorb.) kein LQT signifikant wird und deshalb auch kein Item auszuschließen ist.

Für zukünftige Analysen scheint Aufgrund der neuesten Ergebnisse nach Kubinger et al. (in Vorb.) die Notwendigkeit gegeben, einen neuen *state of the art* zur Prüfung der Modellgültigkeit zu erarbeiten. Aufgrund der dargestellten Ergebnisse rückt weiters die Frage ins Blickfeld des Interesses, wodurch der Unterschied zwischen jenen Items, die anhand inhaltlicher Begründungen ausgeschlossen wurden liegt, im Vergleich zu jenen Aufgaben die teststatistisch auffällig werden. Für Analysen dieser Art bedürfte es allerdings detaillierter Dokumentationen hinsichtlich der inhaltlichen Aspekte, die zum Ausschluss von Items führen. Für eine solche Dokumentation müssten im Vorfeld erst neue Standards erarbeitet werden.

## 7. Zusammenfassung

Zielsetzung der vorliegenden Arbeit war die Klärung der Fragestellung, ob der Ausschluss von einzelnen Aufgaben im Zuge einer vollständigen Überarbeitung und Neugestaltung des Adaptiven Intelligenzdiagnostikums für den AID 3 (Version 3.1, Kubinger & Holocher-Ertl, 2014) basierend auf inhaltlichen Überlegungen retrospektiv auch testtheoretisch fundierbar ist. Ausgangspunkt der retrospektiven Analyse waren Items des AID 2 (Version 2.1, Kubinger & Wurst, 2000 sowie Version 2.2, Kubinger, 2009), welche inhaltlich auffällig waren und aus diesem Grund im AID 3 (Version 3.1; Kubinger & Holocher-Ertl, 2014) ausgeschlossen wurden. Die Operationalisierung der Aufgabenstellung erfolgte mittels einer Modellprüfung unter Anwendung des dichotomen logistischen Modells von *Rasch*.

Das Vorgehen orientierte sich an dem *state of the art* (Kubinger 2005). Dabei gelangte als Modelltest der *Likelihood-Quotienten-Test* nach *Andersen* (1973) mit den Teilungskriterien Score, Geschlecht und Muttersprache zur Anwendung. Es folgte eine Grafische Modellkontrolle nach *Rasch*, auf Basis derer solange schrittweise einzelne Aufgaben ausgeschlossen wurden, bis a-posteriori für die untersuchten Untertests 1, 2, 3, 4, 6, 9, 10 und 11 ein dem dichotomen logistischen Modell nach *Rasch* entsprechender Itempool gewonnen wurde.

Die für die Analyse herangezogene Stichprobe besteht aus 445 Kindern und Jugendlichen im Alter von sechs bis fünfzehn Jahren, welche im Zeitraum von 2001 bis 2013 von zwei erfahrenen klinischen und Gesundheitspsychologinnen einer psychologisch diagnostischen Untersuchung zugeführt wurden.

Sowohl bei Frau Dr. Kaufmann als auch bei Frau Dr. Aumüller wurden die Kinder und Jugendlichen im Rahmen einer Einzelfalldiagnostik zur Klärung einer spezifischen Fragestellung untersucht. Zur Beantwortung der Fragestellungen kamen bei jeder der Testungen neben dem AID 2 (Version 2.1 oder Version 2.2) verschiedene zusätzliche, auf die Fragestellung abgestimmte, psychologisch diagnostische Verfahren zum Einsatz. Die Stichprobe zeichnet sich nicht nur dadurch aus, dass sie von zwei erfahrenen Psychologinnen erhoben wurde, sondern auch dadurch, dass jede Testung aufgrund der individuellen Fragestellung für den Einzelfall bedeutsam war.

Es konnte festgestellt werden, dass im Rahmen der retrospektiven teststatistischen Analyse des AID 2 (Version 2.1 und Version 2.2) weniger Aufgaben auszuschließen waren (insgesamt 34), als dies bei dem, auf inhaltlichen Überlegungen beruhendem Itemausschluss für den AID 3 (Version 3.1) der Fall war. Für den AID 3 wurden auf Grund inhaltlicher Überlegungen 107 Items ausgeschlossen.

Selbst wenn für den AID 2 jene Aufgaben, die bei der teststatistischen Analyse nicht analysierbar waren, zu jenen hinzugenommen wurden, die für den AID 2 auszuschließen waren, konnte das Ergebnis bestätigt werden. Wurden für den AID 2 sowohl die in der teststatistischen Analyse auffällig gewordenen Items als auch die auszuschließenden Aufgaben gemeinsam der Menge von Aufgaben gegenübergestellt, die auf Basis inhaltlicher Überlegungen für den AID 3 ausgeschlossen wurden zeigte sich folgendes Ergebnis: bei vier Untertests 2, 6, 9 und 11 der sechs miteinander Vergleichenen Untertests 1, 2, 3, 6, 9, 11 werden immer noch mehr Items auf Basis inhaltlicher Überlegungen ausgeschlossen, als auf Basis der teststatistischen Analysen für den AID 2 (vgl. Kapitel 6.9). Der inhaltlich Begründete Ausschluss von Aufgaben aus dem AID 2 ist bezüglich des Ausschlusses von Aufgaben strenger, als eine teststatistische Analyse.

Anhand dieser Ergebnisse lässt sich die Notwendigkeit sowohl der inhaltlichen Auseinandersetzung mit den Iteminhalten als auch die teststatistische Analyse der Aufgaben für die Testkonstruktion ableiten. Dies erscheint auch hinsichtlich der Ergebnisse interessant, da beide Vorgehensweisen nicht zu deckungsgleichen Ergebnissen kamen. Für das Adaptive Intelligenzdiagnostikum lässt sich festhalten, dass trotz der Möglichkeit eine Modellpassung durch Verzicht auf eine relativ geringe Menge an Items auf Basis von teststatistischen Analysen erreichen ließe, nicht vom Einsatz einer Itemselektion anhand von inhaltlichen Überlegungen Abstand genommen wird. Dies obwohl durch das Vorgehen weitaus mehr Items ausgeschlossen wurden, als auf Grund der teststatistischen Analysen notwendig gewesen wäre.

Die Modellpassung des AID, welche bereits hinreichend belegt ist (vgl. Kubinger & Wurst, 2000, Kubinger 2009, und Kubinger & Holocher-Ertl 2014), konnte erneut bestätigt werden.

Abschließend ist darauf hinzuweisen, dass aufgrund neuester Ergebnisse (Kubinger, Steinfeld & Reif, in Vorb.) alle *Likelihood-Quotienten-Tests* nach *Andersen* nicht signifikant sind, und daher kein Item auszuschließen ist. Diese neuen Erkenntnisse sind unbedingt bei der Betrachtung der dargestellten Ergebnisse zu beachten.

#### IV. Literaturverzeichnis

- Andersen, E. B. (1973). A Goodness of fit test for the Rasch Model. *Psychometrika*, 38(1), 123-140.
- Fischer, G. H. (1974). Einführung in die Theorie psychologischer Tests. Bern: Huber.
- Kubinger, K. D. (1989a). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie* (S. 19-83). (2., verbesserte Aufl.). Weinheim: Beltz.
- Kubinger, K. D. (1989b). Die Optimierung der Meßgenauigkeit beim "branched"-adaptiven Testen. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie* (S. 187-218) (2., verbesserte Aufl.). Weinheim: Beltz.
- Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenzdiagnostikum* 2. Göttingen: Belz.
- Kubinger, K. D. (2003). Testtheorie, probabilistische. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415-423). Weinheim: Beltz.
- Kubinger, K. D. (2005). Psychological Test Calibration Using the Rasch Model - Some Critical Suggestions on Traditional Approaches, *International Journal of Testing*, 5(4), 377-394, doi: 10.1207/s15327574ijt0504\_3
- Kubinger, K. D. (2006). *Psychologische Diagnostik - Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.

- Kubinger, K. D. (2009). *Adaptives Intelligenz Diagnostikum - Version 2.2 (AID 2) samt AID 2-Türkisch*. Göttingen: Beltz.
- Kubinger, K.D. & Holocher-Ertl, S. (2014). *Adaptives Intelligenz Diagnostikum - Version 3.1 (AID 3)*. Göttingen: Beltz.
- Kubinger, K. D., Steinfeld, J. & Reif, M. (in Vorb.). Item bias and empirical distribution of Andersen's Likelihood-Ratio test in Rasch model-based branched and tailored testing designs.
- Patrick Mair, Reinhold Hatzinger und Marco J. Maier (2014). eRm: Extended Rasch Modeling. R Paketversion 0.15-4. Abgerufen von <http://CRAN.R-project.org/package=eRm>
- R Core Team. (2013). R: A Language and Environment for Statistical Computing (Version 0.15-4) [Software]. The R Foundation for Statistical Computing, Wien. Abgerufen von <http://www.R-project.org/>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. The Danish Institute for Educational Research. Kopenhagen: Pädagogisches Institut.
- Rasch, G. (1980). *Probabilistic Models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.

## V. Anhang

### 8. Abstract - deutsch

Bei der vorliegenden Diplomarbeit handelt es sich um eine retrospektive Analyse der Items der Untertests 1, 2, 3, 4, 6, 9, 10 und 11 des Adaptiven Intelligenzdiagnostikums 2 (AID 2, Version 2.1, Kubinger & Wurst, 2000 und Version 2.2, Kubinger, 2009). Es wurden jene Items, die aus inhaltlichen Gründen nicht in die aktuellste Fassung des AID, den AID 3 (Version 3.1, Kubinger & Holocher-Ertl, 2014) übernommen wurden, mit jenen Aufgaben verglichen, die aus teststatistischen Gründen bei der retrospektiven Analyse des AID 2 auszuschließen waren. Die Modellgültigkeit für das dichotome logistische Modell nach *Rasch* wurde mittels des *Likelihood*-Quotienten-Tests (LQT) nach *Andersen* (1973) anhand der Teilungskriterien Score, Geschlecht und Muttersprache geprüft. Anhand der Grafischen Modellkontrolle nach *Rasch* wurden solange schrittweise Items ausgeschlossen, bis a-posteriori ein *Rasch*-Modell konformer Itempool gewonnen wurde. Bei den in dieser Arbeit untersuchten Untertests waren insgesamt 34 Items des AID 2 (Version 2.1 und Version 2.2) auszuschließen. Im Vergleich dazu wurden für den AID 3 (Version 3.1) in den entsprechenden Untertests insgesamt 107 Items auf Basis inhaltlicher Überlegungen ausgeschlossen. Dieses Ergebnis legt den Schluss nahe, dass der Ausschluss von Items anhand von inhaltlichen Überlegungen, im Gegensatz zum Itemausschluss auf Basis teststatistischer Methoden, zu einer stärkeren Reduktion des Itempools führt. Nach neuesten Erkenntnissen von Kubinger, Steinfeld und Reif (in Vorb.) ist allerdings kein LQT signifikant und deshalb ist auf Basis des gewählten Vorgehens kein einziges Item auszuschließen.

Schlüsselwörter: Rasch-Model; AID 2;

## 9. Abstract - english

The present thesis represents a retrospective analysis of items of the subtests 1, 2, 3, 4, 6, 9, 10 and 11 of the Adaptiven Intelligenzdiagnosticums (AID 2, version 2.1, Kubinger & Wurst, 2000 as well as version 2.2, Kubinger, 2009). All items which needed to be excluded in the newest generation of the AID (AID 3, version 3.1, Kubinger & Holoher-Ertl, 2014), based on contentual reconsiderations, were compared with those items, that should be precluded due to teststatistical analysis of the AID 2. To verify the conformance of the itempool to the 1-PL model, Likelihood-Ratio tests (LRTs, Andersen, 1973) were applied using the splitting criteria score, sex and first language. Based on the graphical model-check of *Rasch*, items had been excluded until the itempool fitted a-posteriori the 1-PL model. Based on this analysis 34 items had to be excluded from the subtests of the AID 2. On the Contrary, 107 items were excluded for the new AID 3 (version 3.1) in regard to the content. The conclusion is drawn that the elimination of items based on content, led in contrast to the exclusion of items by using teststatistical methods, to a far more reduced itempool. It needs to be mentioned that according to recent research by Kubinger, Steinfeld und Reif (in press) none of the conducted LRTs are significant and so, on basis of the chosen methodological approach, no item at all needs to be excluded. Future research should therefore deal with establishing a new state of the art method for conducting the teststatistical approach.

Keywords: 1-PL model; AID 2;

## 10. Curriculum Vitae

### Persönliche Informationen

---

**Name:** Urban Lorenz

### Tabellarischer Lebenslauf

---

<b>1990-1998</b>	BG/BRG Stockerau Fachbereichsarbeit in Psychologie: „Das Depressive Syndrom“
<b>1998-1999</b>	Flüchtlingsbetreuungsstelle der Caritas St. Gabriel Maria Enzersdorf
<b>1999-2002</b>	Medizinstudium an der Universität Wien
<b>seit 2002</b>	Psychologiestudium an der Universität Wien

### Tätigkeitsübersicht

---

<b>2006</b>	Therapiezentrum Ybbs, Prim. I Bereich: psychotherapeutische Behandlung von Alkohol / Suchtkranken; Praktikum im Umfang von 160 Stunden
<b>2010</b>	Landeskrankenhaus Mostviertel Amstetten-Mauer 4. Abteilung; Bereich für Forensische Psychiatrie 6-Wochen-Praktikum im Umfang von 240 Stunden
<b>2010-2011</b>	Österreichische Gesellschaft für Familienplanung Weiterbildungskursvorbereitung, Evaluation, Projektarbeit;