

►Exkurs: Alpha-Adjustierung

Werden mehrfach Tests an einem Datensatz durchgeführt (multiples Testen), so besteht die Gefahr der Alpha-Kumulierung. Dies bedeutet das Risiko fälschlicher Weise die Modellgültigkeit zu verwerfen (H_0 , Fehler 1. Art) wird über die Anzahl der Tests immer größer. Es beträgt $1 - (1 - \alpha)^q$, wobei q die Anzahl der Tests bezeichnet. Bei Alpha .05 würde die Irrtumswahrscheinlichkeit bei sieben Tests auf $1 - (.95)^7 = .30$ ansteigen (30%). Um dem entgegen zu wirken wird für die einzelnen Tests so korrigiert, dass das gemeinsame Alpha für alle Tests dem festgesetzten Wert entspricht. Nach der Bonferroni-Korrektur wird das Alpha durch die Anzahl der Tests dividiert ($\alpha/q = \alpha^*$). Da die Wahrscheinlichkeit die Modellgültigkeit fälschlicherweise anzunehmen gering sein soll (Fehler 2. Art) wird das Alpha auf 5% festgesetzt ($\alpha = .05$). Für die einzelnen Tests beträgt das korrigierte Alpha nun $0.05/7 = 0.007$ (siehe Bortz & Döring, 2006; Koller, Alexandrowicz, Hatzinger, 2012).

Da mit steigender Anzahl an Tests das Alpha stetig sinkt, werden nur inhaltlich relevanten Teilungskriterien verwendet. Dies kann dennoch bei einer Erstanalyse deutlich von Nachteil sein, deshalb wird hier für alle Tests Alpha auf .01 festgesetzt.

(c) Item-Analyse & Wald-Test

Bei signifikantem LQT wird zur genaueren Item-Analyse ein Wald-Test durchgeführt. Bei diesem Test werden die geschätzten Parameter und Standardfehler für jedes Item einzeln untersucht, um auffällige Items leichter identifizieren zu können.

Prinzipiell stellt ein theoriegeleitetes Vorgehen die Methode der Wahl dar (a priori Hypothesen über Gründe von Modellabweichungen). Items, welche mit dem Rasch-Modell nicht verträglich sind, sollen prinzipiell aufgrund theoretisch-inhaltlicher Überlegungen selektiert werden, um eine artifizielle (künstliche) Modellanpassung zu vermeiden (Gittler, 1986). Hierbei dient der Wald-Test lediglich als zusätzliche Hilfestellung um, falls notwendig, mögliche a posteriori-Hypothesen für eine Item-Selektion formulieren zu können.

(d) Martin-Löf-Test (1973)

Zur Überprüfung der Itemhomogenität von Subskalen wird der LQT nach Martin Löf (1973) verwendet. Anstatt verschiedener Personengruppen werden hier unterschiedliche Itemgruppen betrachtet. Sowohl innerhalb der Subtests des TeMuK,

als auch über mehrere Skalen hinweg wird überprüft, ob sie eine gemeinsame (übergeordnete) latente Dimension erfassen und somit Eindimensionalität für die Items angenommen werden kann. Innerhalb eines Subtests wurden folgende Teilungskriterien herangezogen:

- der Median der Itemscores, d. h. leichte vs. schwere Items,
- die Art der Aufgabenstellung (z.B. höher vs. tiefer)
- die Itemreihenfolge, d. h. Items am Anfang des Tests vs. Items am Testende (Ermüdung)

9.1. Modellprüfung: LQT nach Anderson (1973)

‣ Subskala Tonhöhe

Die erste Rasch-Modellprüfung mittels LQT wurde für die Subskala Tonhöhe unter den sieben Teilungskriterien durchgeführt. Einige Items wiesen beim Teilungskriterium Median und Mittelwert unangemessene Antwortmuster auf, d.h. nicht jede Antwortkategorie kam in jeder Subgruppe vor. Diese Items mussten daher bei den entsprechenden Berechnungen ausgeschlossen werden. Der erste Modelltest für die Subskala Tonhöhe fiel unter den Teilungskriterien Mittelwert, Musikerstatus und Unterrichtsjahre signifikant aus. Das Rasch-Modell gilt somit nicht. Die entsprechenden Kennwerte sind aus *Tabelle 9.1* zu entnehmen.

Bereits a priori galt die Vermutung, dass die wechselnden Tonhöhen der Items die Güte der Skala beeinflussen könnten. D. h. jedes Item des ersten Testteils, worunter sich auch die Items der Skala Tonhöhe befinden, wurde auf einer anderen Tonhöhe vorgegeben, um den Testpersonen die Bearbeitung der Items angenehmer zu gestalten und aversives oder demotiviertes Testverhalten aufgrund zu großer Eintönigkeit zu minimieren (Verfälschung). Werden bspw. die drei Töne eines Items auf einer relativ höheren Tonhöhe präsentiert als jene des nachfolgenden Items, bei welchem die Art der Veränderung der Tonhöhe gemäß der Aufgabe als relativ höher oder tiefer in Relation zu den drei Tönen des *aktuellen* Items und nicht zum vorhergehenden Item beurteilt werden soll, so könnte diese Itemabfolge die Testreaktion beeinflussen. Letztendlich galt die Vermutung, dass durch diesen Umstand die lokalen stochastischen Unabhängigkeit verletzt sein würde. Dies zeigte sich auch in der folgenden Item-Analyse mittels Wald-Test. Es fielen Item 15 und 45 mit Signifikanz auf. Diese Items lagen einen Ton tiefer als ihr vorhergehendes Item und

die Art der Veränderung war zusätzlich nur 1 bis 2 Herz tonabwärts gerichtet (schwierigste Items). Im zweiten Durchgang der Modelltests wurde das schwierigere Item 15 ausgeschlossen, im dritten Item 45. Nach neuerlicher Signifikanz wurde auch Item 27 entfernt, es stellt den umgekehrten Fall dar: Das Item lag einen Ton höher als das vorhergehende und die Art der Veränderung war zusätzlich 1 Herz tonaufwärts gerichtet. Für die Subskala Tonhöhe gilt nach dem vierten Testdurchgang das Rasch-Modell. Die Kennwerte des vierten und letzten Durchgangs an Modelltests sind in *Tabelle 9.2* abgebildet.

Tabelle 9.1: Ergebnisse des 1. Durchgang der Modelltests (Andersen, 1973) für die Subskala Tonhöhe bei $N = 174$, $k = 16$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	100	hoch	74	3,5,13,25,37	18.553	10	23.2093	.046
Mittelwert	niedrig	73	hoch	101	3,5,25	74.672	12	26.217	<.001*
Geschlecht	männlich	81	weiblich	93	-	11.259	15	30.5779	.734
Alter	jünger	82	älter	92	-	12.235			.661
Bildungs-stand	ohne Matura	91	mit Matura	83	-	15.863			.391
Musiker-status	Nicht- u. Hobby-M.	112	Semi- bis Profi-M.	62	-	44.665			<.001*
Unterrichts-jahre	<=6	70	>6	104	-	40.745			<.001*

Tabelle 9.2: Ergebnisse des letzten Durchgang des Modelltests (Andersen, 1973) für die Subskala Tonhöhe nach Ausschluss der Items 15, 45, 27 bei $N = 174$, $k = 13$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	88	hoch	86	3,5,13,25,37	11.143	7	18.4753	.132
Mittelwert	niedrig	67	hoch	107	3,5	18.826	10	32.2093	.043
Geschlecht	männlich	81	weiblich	93	-	12.086	12	26.217	.439
Alter	jünger	82	älter	92	-	13.806			.313
Bildungs-stand	ohne Matura	91	mit Matura	83	-	11.888			.455
Musiker-status	Nicht- u. Hobby-M.	112	Semi- bis Profi-M.	62	-	11.86			.457
Unterrichts-jahre	<=6	70	>6	104	-	23.4			.025

► Subskala Lautstärke

Die Items der Subskala Lautstärke erweisen sich unter allen Teilungskriterien bereits beim ersten Durchgang als Rasch-homogen (*Tabelle 9.3*).

Tabelle 9.3: Ergebnisse des Modelltests (Andersen, 1973) für die Subskala Lautstärke bei $N = 174$, $k = 15$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	101	hoch	73	11	24.042	13	27.6883	.031
Mittelwert	niedrig	84	hoch	90	-	25.756	14	29.1412	.028
Geschlecht	männlich	81	weiblich	93	-	14.167			.437
Alter	jünger	82	älter	92	-	10.132			.752
Bildungs-stand	ohne Matura	91	mit Matura	83	-	24.817			.036
Musiker-status	Nicht- u. Hobby-M.	112	Semi- bis Profi-M.	62	-	18.49			.185
Unterrichts-jahre	<=6	70	>6	104	-	11.149			.674

► Subskala Zeitmaß

Für die Subskala Zeitmaß gilt das Rasch-Modell (siehe *Tabelle 9.4*).

Tabelle 9.4: Ergebnisse des Modelltests (Andersen, 1973) für die Subskala Zeitmaß bei $N = 174$, $k = 16$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	114	hoch	60	1,6,10, 16,28, 29,31	3.888	8	20.0902	.867
Mittelwert	niedrig	78	hoch	96	29	11.154	14	29.1412	.674
Geschlecht	männlich	81	weiblich	93	-	24.713	15	30.5779	.054
Alter	jünger	82	älter	92	-	14.191			.511
Bildungs-stand	ohne Matura	91	mit Matura	83	-	11.329			.729
Musiker-status	Nicht- u. Hobby-M.	112	Semi- bis Profi-M.	62	-	16.7			.337
Unterrichts-jahre	<=6	70	>6	104	-	14.311			.502

► Subskala Tongedächtnis

Der Modelltest für die Subskala Tongedächtnis wurde im ersten Durchgang bei den Teilungskriterien Median und Mittelwert sowie Musikerstatus signifikant (Tabelle 9.5). In den folgenden zwei Durchgängen wurden zuerst Item 6 und 11 nacheinander ausgeschieden. Diese Items wurden bereits a priori als besonders schwierig bzw. kaum lösbar eingestuft, da der geänderte Ton dieselbe musikalische Funktion wie jener der Original-Melodie hat. Diese Items zeigten sich auch bei der folgenden Analyse im Wald-Test auffällig (signifikant). Item 7, welches dem Motiv einer bekannten Schlusswendung gleicht, zeigte sich schließlich für das Teilungskriterium Geschlecht auffällig und wurde im vierten Durchgang ausgeschieden. Nach Ausschluss dieser drei Items gilt für die Subskala Tongedächtnis das Rasch-Modell (Tabelle 9.6).

Tabelle 9.5: Ergebnisse des 1. Durchgang der Modelltests (Andersen, 1973) für die Subskala Tongedächtnis bei $N = 174$, $k = 17$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	93	hoch	81	1,4,10	44.375	13	27.6882	< .001*
Mittelwert	niedrig	81	hoch	93	1,10	53.089	14	29.1412	< .001*
Geschlecht	männlich	81	weiblich	93	-	31.199	16	31.9999	.013
Alter	jünger	82	älter	92	-	23.626			.098
Bildungs-stand	ohne Matura	91	mit Matura	83	-	14.931			.530
Musiker-status	Nicht- u. Hobby-M.	112	Sem- bis Profi-M.	62	-	40.578			.001*
Unterrichts-jahre	<=6	70	>6	104	-	26.858			.043

Tabelle 9.6: Ergebnisse des letzten Durchgang des Modelltests (Andersen, 1973) für die Subskala Tongedächtnis nach Ausschluss der Items 6, 11, 7 bei $N = 174$, $k = 14$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	101	hoch	73	1,4,10,17	19.635	9	21.666	.020
Mittelwert	niedrig	81	hoch	93	1	22.981	12	26.217	.028
Geschlecht	männlich	81	weiblich	93	-	18.357	13	27.6882	.144
Alter	jünger	82	älter	92	-	18.659			.134
Bildungs-stand	ohne Matura	91	mit Matura	83	-	14.592			.333
Musiker-status	Nicht- u. Hobby-M.	112	Sem- bis Profi-M.	62	-	20.515			.083
Unterrichts-jahre	≤ 6	70	> 6	104	-	14.544			.337

► Subskala Rhythmusgedächtnis

Die Items der Subskala Rhythmusgedächtnis erweisen sich als Rasch-homogen (Tabelle 9.7).

Tabelle 9.7: Ergebnisse des der Modelltests (Andersen, 1973) für die Subskala Rhythmusgedächtnis bei $N = 174$, $k = 16$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	108	hoch	66	-	10.584	15	30.5779	.781
Mittelwert	niedrig	73	hoch	101	-	16.685			.338
Geschlecht	männlich	81	weiblich	93	-	13.639			.553
Alter	jünger	82	älter	92	-	15.363			.426
Bildungs-stand	ohne Matura	91	mit Matura	83	-	24.921			.051
Musiker-status	Nicht- u. Hobby-M.	112	Sem- bis Profi-M.	62	-	25.971			.038
Unterrichts-jahre	≤ 6	70	> 6	104	-	25.357			.045

►Skalen-Kombinationen

Für die Subskalen des ersten Testteils Wahrnehmung gilt das Rasch-Modell nicht. Die zwei Skalen Zeitmaß und Lautstärke erweisen sich jedoch zusammen als Rasch-homogen (Tabelle 9.8). Die Skalen Tonhöhe und Tongedächtnis zeigen sich im Teilungskriterium Mittelwert als nicht Rasch-homogen. Für die Skalen Zeitmaß und Rhythmusgedächtnis gilt jedoch das Raschmodell (Tabelle 9.9). Wo die Skala Zeitmaß endgültig zuzuordnen ist kann an dieser Stelle nicht eindeutig beantwortet werden. Für eine Diskussion sei nachfolgend der Martin-Löf-Tests auf Kap. 9.2 verwiesen.

Tabelle 9.8: Ergebnisse des der Modelltests (Andersen, 1973) für die Subskalen Zeitmaß und Lautstärke zusammen bei $N = 174$, $k = 31$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	104	hoch	70	12,18,29	31.754	27	46.963	.241
Mittelwert	niedrig	70	hoch	104	-	50.831	30	50.8922	.010
Geschlecht	männlich	81	weiblich	93	-	46.035			.031
Alter	jünger	82	älter	92	-	27.706			.586
Bildungs-stand	ohne Matura	91	mit Matura	83	-	34.652			.256
Musiker-status	Nicht- u. Hobby-M.	112	Sem- bis Profi-M.	62	-	31.657			.384
Unterrichts-jahre	≤ 6	70	> 6	104	-	20.926			.890

Tabelle 9.9: Ergebnisse des der Modelltests (Andersen, 1973) für die Subskalen Zeitmaß und Rhythmusgedächtnis zusammen bei $N = 174$, $k = 32$, $\alpha = .01$

Teilungs-kriterien	Gruppe 1	n_1	Gruppe 2	n_2	ohne Item	χ^2 empirisch	df	χ^2 kritisch ($\alpha = .01$)	p
Median	niedrig	90	hoch	84	29 (Skala Zeitmaß)	25.405	30	50.8922	.705
Mittelwert	niedrig	77	hoch	97		27.896			.576
Geschlecht	männlich	81	weiblich	93	-	47.476	31	52.1913	.030
Alter	jünger	82	älter	92	-	32.655			.386
Bildungs-stand	ohne Matura	91	mit Matura	83	-	36.147			.241
Musiker-status	Nicht- u. Hobby-M.	112	Sem- bis Profi-M.	62	-	46.099			.040
Unterrichts-jahre	≤ 6	70	> 6	104	-	45.469			.045

9.2. Test nach Martin LÖF (1973)

Bei Überprüfung der Item-Homogenität für die Subskala Tonhöhe zeigte sich nach den Teilungskriterien Median und Itemreihenfolge (Items am Anfang vs. Ende) keine Signifikanz. Bei Betrachtung der Art der Aufgabenstellung (Veränderung des Tones in Richtung höher vs. tiefer) kann jedoch keine Eindimensionalität der Subskala Tonhöhe angenommen werden. (siehe *Tabelle 9.10*). In den folgenden *Tabellen 9.11 bis 9.16* sind alle Martin-Löf-Tests zur Überprüfung der Item-Homogenität für alle Subskalen und Skalen-Kombinationen angeführt. Keiner dieser Tests wurde signifikant, d.h. es konnte keine Verletzung der Item-Homogenität bzw. Eindimensionalität festgestellt werden. Verschiedene Testteile (z. B. Median: schwierige vs. leichte) erfassen dieselbe latente Dimension. Auch am Testende fließen keine nennenswerten Ermüdungseffekte in unterschiedlichem Maße ein (Testanfang vs. Testende).

Tabelle 9.10: Ergebnisse des Test nach Martin Löf (1973) für die Subskala Tonhöhe nach Ausschluss der Items 15, 45 und 27 (k=13)

Teilungskriterium	Gruppe 1	k ₁	Gruppe 2	k ₂	χ^2 empirisch	χ^2 kritisch ($\alpha = .01$)	df	p
Median	leicht	6	schwer	7	33.273	64.9501	41	.799
Aufgabenstellung	höher	7	tiefer	6	77.514			<.001*
Itemreihenfolge	Anfang	7	Ende	6	37.197			.640

Tabelle 9.11: Ergebnisse des Martin-Löf-Tests (1973) für die Subskala Lautstärke (k=15)

Teilungskriterium	Gruppe 1	k ₁	Gruppe 2	k ₂	χ^2 empirisch	χ^2 kritisch ($\alpha = .01$)	df	p
Median	leicht	7	schwer	8	56.964	82.2921	55	.402
Aufgabenstellung	lauter	7	leiser	8	76.936			.027
Itemreihenfolge	Anfang	8	Ende	7	44.846			.834

Tabelle 9.12: Ergebnisse des Test nach Martin Löff (1973) für die Subskala Zeitmaß (k=16)

Teilungskriterium	Gruppe 1	k ₁	Gruppe 2	k ₂	χ^2 empirisch	χ^2 kritisch ($\alpha = .01$)	df	p
Median	leicht	8	schwer	8	29.8	92.01	63	>.999
Aufgabenstellung	vorgezogen	8	verzögert	8	80.297			.070
Itemreihenfolge	Anfang	8	Ende	8	52.091			.835

Tabelle 9.13: Ergebnisse des Test nach Martin Löff (1973) für die Subskala Tongedächtnis nach Ausschluss der Items 6, 11 und 7 (k=14)

Teilungskriterium	Gruppe 1	k ₁	Gruppe 2	k ₂	χ^2 empirisch	χ^2 kritisch ($\alpha = .01$)	df	p
Median	leicht	7	schwer	7	32.3	117.0565	48	.960
Aufgabenstellung	höher	8	tiefer	6	39.454			.775
Itemreihenfolge	Anfang	7	Ende	6	48.675			.446

Tabelle 9.14: Ergebnisse des Test nach Martin Löff (1973) für die Subskala Rhythmusgedächtnis (k=16)

Teilungskriterium	Gruppe 1	k ₁	Gruppe 2	k ₂	χ^2 empirisch	χ^2 kritisch ($\alpha = .01$)	df	p
Median	leicht	8	schwer	8	55.526	92.01	63	.737
Aufgabenstellung	erster Rhythmus richtig	7	erster Rhythmus nicht richtig	9	37.608	90.8015	62	.994
Itemreihenfolge	Anfang	7	Ende	8	58.963	92.01	63	.621

Tabelle 9.15: Ergebnisse des Test nach Martin Löff (1973) für die Subskalen Lautstärke und Zeitmaß zusammen (k=31)

Teilungskriterium	Gruppe 1	k ₁	Gruppe 2	k ₂	χ^2 empirisch	χ^2 kritisch ($\alpha = .01$)	df	p
Median	leicht	15	schwer	16	95.275	292.7818	23 9	>.999
Aufgabenstellung	Zeitmaß	16	Lautstärke	15	240.878		23 9	.454

Tabelle 9.16: Ergebnisse des Test nach Martin Löff (1973) für die Subskalen Zeitmaß und Rhythmusgedächtnis zusammen (k=32)

Teilungskriterium	Gruppe 1	k ₁	Gruppe 2	k ₂	χ^2 empirisch	χ^2 kritisch ($\alpha = .01$)	df	p
Median	leicht	16	schwer	16	127.43	310.4574	255	>.999
Aufgabenstellung	Rhythmus-Gedächtnis	16	Zeitmaß	16	203.496			.992

Laut den Modelltests nach Andersen (1973) und den Martin-Löff-Tests messen die Skalen Lautstärke und Zeitmaß eine gemeinsame latente Fähigkeit sowie die Skalen Zeitmaß und Rhythmusgedächtnis zusammen. Abgesehen davon, dass die Frage der die Einordnung der Skala Zeitmaß sowie der übrigen Skalen offen bleiben muss, sollten die Ergebnisse kritisch betrachtet werden. Stelzel (1979) kritisierte bereits, dass bei vorliegen zweier Rasch-homogener Subskalen in einem zusammengesetzten Test die Modelltests nicht sensitiv genug sind. Gittler (1999) geht diesem Problem bereits nach und zeigt auf wie dieses Problem mittels spezielle Teilungsgruppenbildung im Andersen-Test (1973) und anderer trennscharfer Tests gelöst werden kann. Weitere Analysen des TeMuK sollten sich mit einer genügend großen Stichprobe daran orientieren. Vorerst wurden die Skalen einzeln verrechnet. das Problem nur angedacht.

9.3. Person-Item-Map

Im oberen Teil der Person-Item-Map ist die Verteilung der Personenparameter dargestellt. Sie liefert einen Eindruck über das Fähigkeitsniveau der Stichprobe. Darunter ist die Position der Items auf der latenten Dimension abgebildet. Verhältnismäßig leichtere Items befinden sich weiter links und schwerere Items befinden sich weiter rechts in der Grafik (bzw. auf der latenten Dimension). Anhand der Punktwolke kann das Schwierigkeitsniveau des Tests beurteilt werden. Im Idealfall sind die Punkte (Itemschwierigkeiten) über den ganzen Bereich (der latenten Dimension) verteilt um die latente Dimension möglichst gut und differenziert zu erfassen.

Aus *Abbildung 9.1 bis 9.5* ist ersichtlich, dass sich die Verteilung der Personenparameter eher auf der rechten Seite häuft, d.h. viele Personen haben einen hohen Score erreicht (siehe *Tabelle 9.17*). Die Items erfassen den mittleren Bereich der latenten Dimension sehr gut. Es gibt jedoch sehr wenige schwierige Items die das

höhere Fähigkeitsniveau ausreichend differenziert erfassen. Genauso sind wenige leichte Items vorhanden, was im Fall dieser Stichprobe weniger ins Gewicht fällt, denn der Test ist eher zu leicht als zu schwer für die Untersuchungsgruppe. Unter diesen Umständen könnte man als repräsentative Stichprobe ohne Überrepräsentation von erfahrenen Musik-Studenten und Profi-Musikern die Gruppe der Nicht-Musiker ($n = 35$) und Hobby-Musiker ($n = 77$) im Bezug auf die Verteilung der Personenparameter in den einzelnen Skalen betrachten. Aus *Abbildung 9.6.* ist ersichtlich, dass sich dabei eine Normalverteilung um den Mittelwert zeigt. D. h. der Test wäre unter Umständen für eine repräsentative Stichprobe zur Untersuchung von Leistungsunterschieden sehr gut geeignet. Vorerst wurde dieser Schritt nur angedacht und sollte in nachfolgenden Analysen mit adäquater Stichprobe genauer betrachtet werden.

Die Benennung der Items am linken Rand der *Abbildungen 9.1 bis 9.5* stellt eine Arbeitsbezeichnung dar.

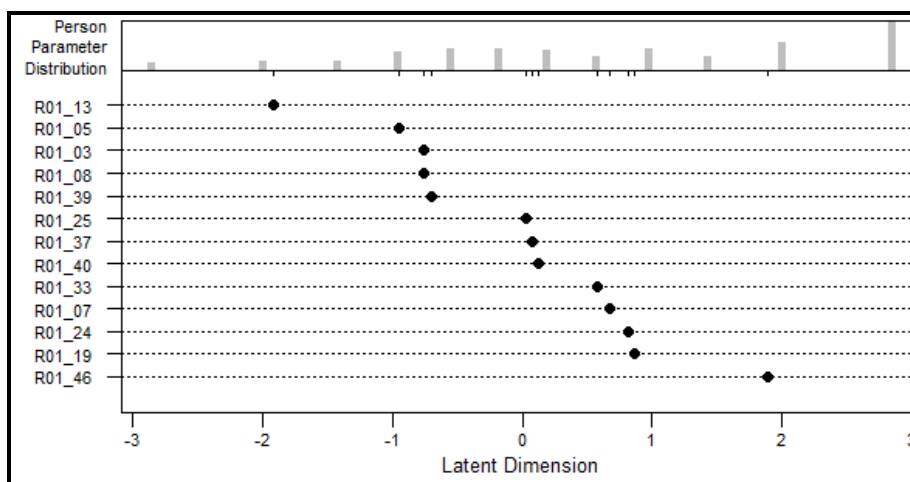


Abbildung 9.1: Person-Item-Map für die Subskala Tonhöhe (k=13)

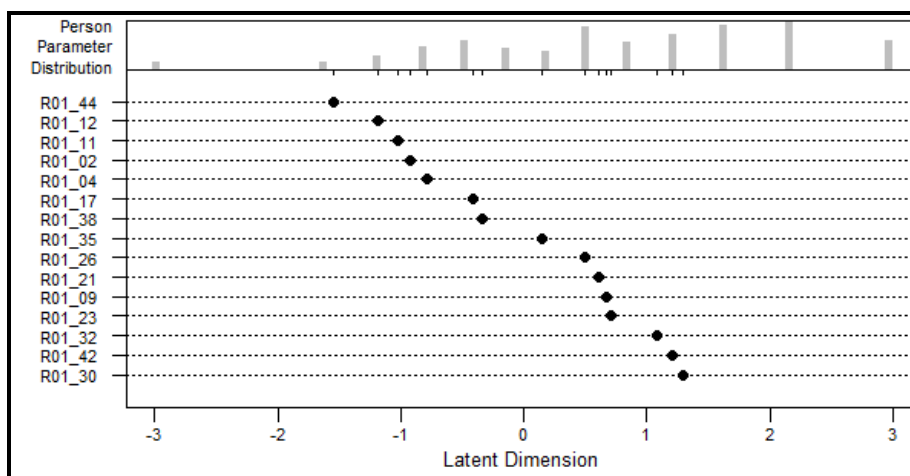


Abbildung 9.2: Person-Item-Map für die Subskala Lautstärke (k=15)

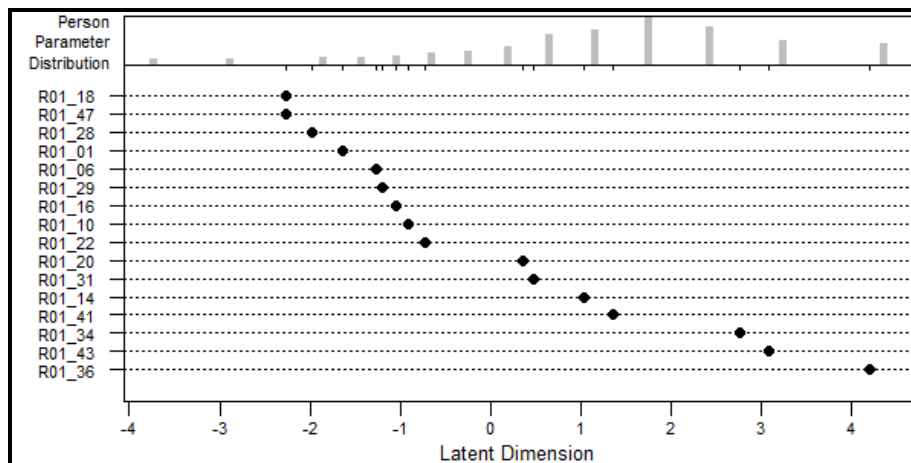


Abbildung 9.3: Person-Item-Map für die Subskala Zeitmaß (k=16)

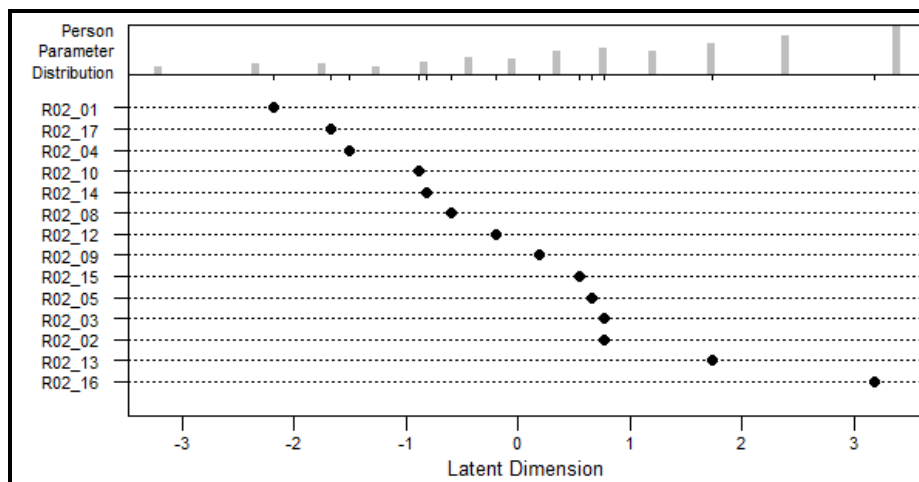


Abbildung 9.4: Person-Item-Map für die Subskala Tongedächtnis (k=17)

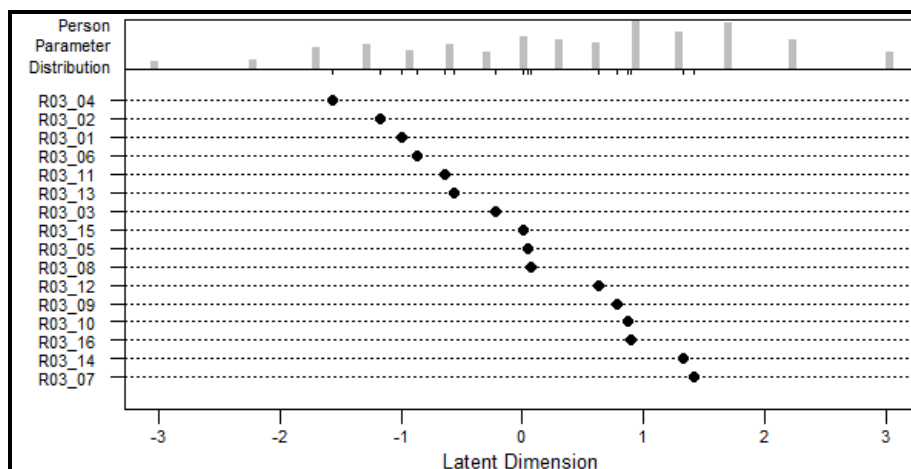


Abbildung 9.5: Person-Item-Map für die Subskala Rhythmusgedächtnis (k=16)

Tabelle 9. 17: Häufigkeiten der Parameterwerte in der Gesamtstichprobe (N = 174)

Personenparameter Tonhöhe	n	Personenparameter Lautstärke	n
-3.7804243	1	-3.7480784	1
-2.8542136	2	-2.9860475	1
-1.9972998	3	-1.6273544	1
-1.4254866	3	-1.1979376	5
-0.9654503	9	-0.8239979	10
-0.5602483	11	-0.4805262	14
-0.1818514	11	-0.1530098	9
0.1880524	10	0.1689008	8
0.5653343	6	0.4944505	22
0.9684377	11	0.8341094	13
1.4256212	6	1.2026082	17
1.9941304	15	1.6252121	23
2.8476309	30	2.1564345	26
3.770421	56	2.9671713	14
		3.8459154	10
Personenparameter Zeitmaß	n	Personenparameter Tongedächtnis	n
-4.6283684	1	-4.15516826	1
-3.7304123	1	-3.21610234	2
-2.8879586	1	-2.34511434	4
-1.8571741	2	-1.75398072	4
-1.4407234	2	-1.27175723	2
-1.0440363	3	-0.84348855	6
-0.6497295	6	-0.44286813	9
-0.2444967	7	-0.05326794	8
0.1834657	11	0.33907427	14
0.6462451	20	0.74914312	17
1.1586196	24	1.19757383	14
1.7411619	36	1.71887801	20
2.4215389	27	2.38275634	27
3.2435327	15	3.38114711	36
4.3568516	13	4.46016084	10
5.5396413	5		
Personenparameter Rhythmusged	n		
-3.913397848	1		
-3.036350359	1		
-2.227884417	2		
-1.701687939	9		
-1.287063043	10		
-0.929748879	7		
-0.604712919	10		
-0.297684764	6		
0.001237493	14		
0.300093539	13		
0.606903494	11		
0.931519981	24		
1.288131289	17		
1.701664572	22		
2.226261898	13		
3.032491328	6		
3.907161715	8		

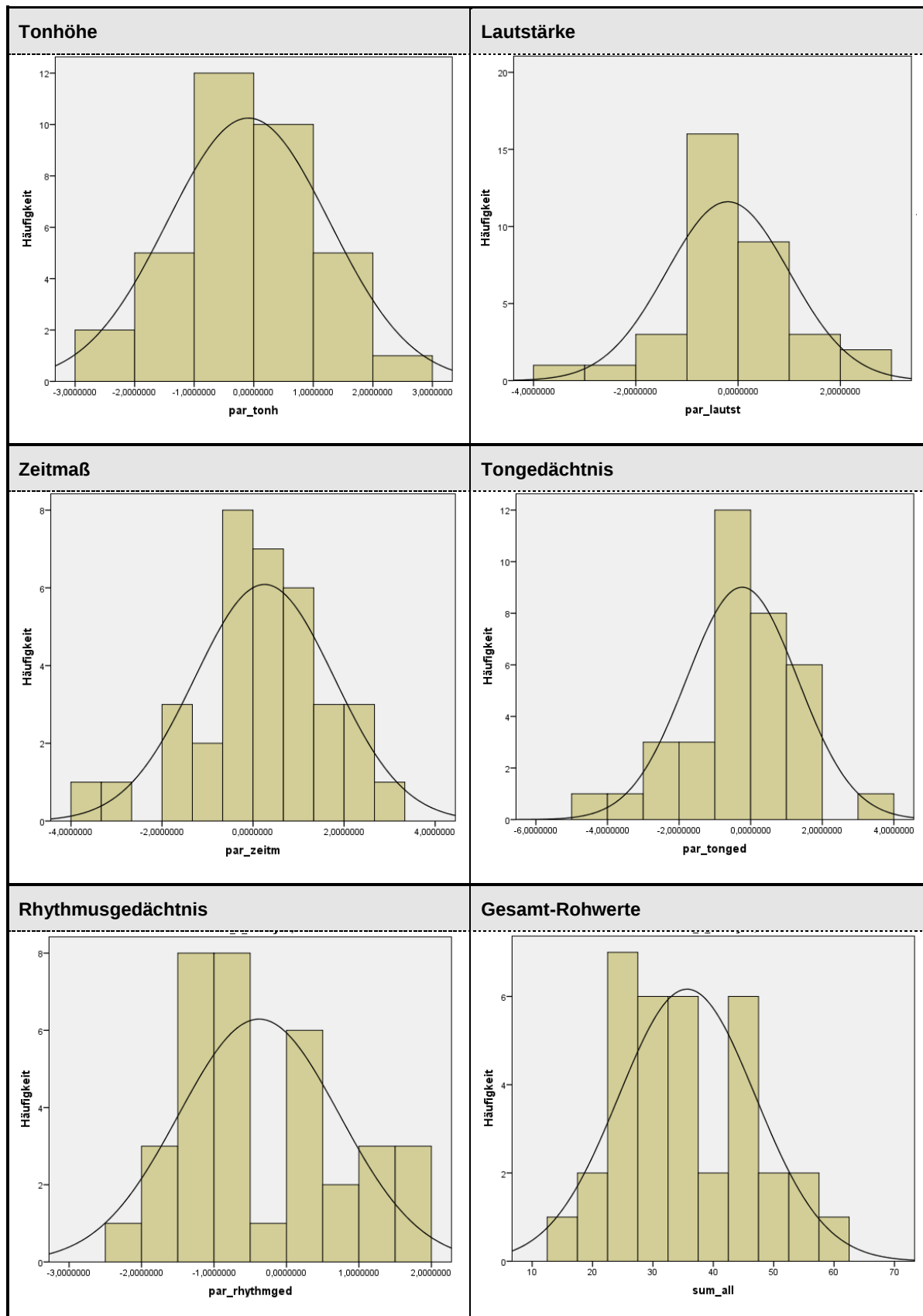


Abbildung 9.6: Verteilung der Personenparameter der einzelnen Skalen sowie des Gesamt-Rohwerts für die Gruppe der Nicht-Musiker ($n = 35$) und Hobby-Musiker ($n = 77$)

9.4. Personenparameter

In *Abbildung 9.9* (linke Spalte) sind für alle Personenscores (Anzahl gelöster Aufgaben) die entsprechenden Personenparameter für jede Subskala aufgelistet. Die Extrembereiche können nicht geschätzt werden, da nicht bekannt ist, wie viel bzw. wenig fähiger jene Personen sind, welche alle bzw. keine Items gelöst haben. Dasselbe gilt für Scores, die keine Person erreicht hat. In der Software werden die Werte für die Extrembereiche interpoliert, daher wird für diese auch kein Standardfehler ausgegeben. Die Schätzungen für die Personenparameter und die Itemparameter sind Summe-Null-normiert, d.h. die Summe der Parameterschätzungen ist jeweils Null. Ein negativer großer Wert für den Personenparameter deutet auf eine geringere Personenfähigkeit hin. Ein großer positiver Wert deutet auf eine hohe Personenfähigkeit hin. In der rechten Spalte (*Grafik 9.9*) sind alle errechneten Item-Leichtigkeits-Parameter aufgelistet. Je größer der Wert im positiven Bereich ist, desto leichter das Item. Für den Standardfehler (*SE*) gilt, je kleiner der Wert, desto genauer die Schätzung.

Tabelle 9.18: Personenparameter und Item-Leichtigkeitsparameter aller Skalen sowie deren Standardfehler für die untersuchte Stichprobe $N = 174$

Tonhöhe: Personenparameter ($k = 16$)			Tonhöhe: Item-Leichtigkeits-Parameter ($k = 16$)		
Raw Score	Estimate	Std. Error		Estimate	Std. Error
0	-3.7804243	NA	beta R01_03	0.754	0.239
1	-2.8542136	1.0786310	beta R01_05	0.944	0.246
2	-1.9972998	0.8151532	beta R01_07	-0.674	0.215
3	-1.4254866	0.7085706	beta R01_08	0.754	0.239
4	-0.9654503	0.6531033	beta R01_13	1.912	0.311
5	-0.5602483	0.6231068	beta R01_19	-0.868	0.214
6	-0.1818514	0.6094835	beta R01_24	-0.819	0.215
7	0.1880524	0.6090199	beta R01_25	-0.024	0.220
8	0.5653343	0.6218009	beta R01_33	-0.576	0.215
9	0.9684377	0.6511888	beta R01_37	-0.075	0.219
10	1.4256212	0.7063937	beta R01_39	0.693	0.237
11	1.9941304	0.8131207	beta R01_40	-0.126	0.218
12	2.8476309	1.0772249	beta R01_46	-1.895	0.219
13	3.7704210	NA			
Lautstärke: Personenparameter ($k = 15$)			Lautstärke: Item-Leichtigkeits-Parameter ($k = 15$)		
Raw Score	Estimate	Std. Error		Estimate	Std. Error
0	-3.7480784	NA	beta R01_02	0.927	0.211
1	-2.9860475	1.0596814	beta R01_04	0.788	0.204
3	-1.6273544	0.6862027	beta R01_09	-0.675	0.170
4	-1.1979376	0.6293406	beta R01_11	1.025	0.216
5	-0.8239979	0.5964148	beta R01_12	1.183	0.225
6	-0.4805262	0.5775369	beta R01_17	0.416	0.190
7	-0.1530098	0.5684691	beta R01_21	-0.615	0.170
8	0.1689008	0.5675960	beta R01_23	-0.705	0.169
9	0.4944505	0.5750121	beta R01_26	-0.494	0.171
10	0.8341094	0.5925015	beta R01_30	-1.295	0.170
11	1.2026082	0.6244420	beta R01_32	-1.089	0.169
12	1.6252121	0.6808216	beta R01_35	-0.147	0.176
13	2.1564345	0.7884108	beta R01_38	0.340	0.187
14	2.9671713	1.0554959	beta R01_42	-1.207	0.169
15	3.8459154	NA	beta R01_44	1.549	0.251

Fortsetzung Tabelle 9.17

Zeitmaß: Personenparameter (k = 16)			Zeitmaß: Item-Leichtigkeits-Parameter (k = 16)		
Raw Score	Estimate	Std. Error		Estimate	Std. Error
0	-4.6283684	NA	beta R01_01	1.636	0.309
1	-3.7304123	1.0694119	beta R01_06	1.273	0.277
2	-2.8879586	0.8094036	beta R01_10	0.909	0.250
4	-1.8571741	0.6588680	beta R01_14	-1.031	0.182
5	-1.4407234	0.6349222	beta R01_16	1.045	0.260
6	-1.0440363	0.6269380	beta R01_18	2.263	0.381
7	-0.6497295	0.6306704	beta R01_20	-0.364	0.193
8	-0.2444967	0.6439552	beta R01_22	0.723	0.239
9	0.1834657	0.6657500	beta R01_28	1.979	0.346
10	0.6462451	0.6962328	beta R01_29	1.193	0.271
11	1.1586196	0.7372128	beta R01_31	-0.472	0.191
12	1.7411619	0.7913493	beta R01_34	-2.765	0.204
13	2.4215389	0.8607861	beta R01_36	-4.201	0.285
14	3.2435327	0.9601533	beta R01_41	-1.365	0.180
15	4.3568516	1.1897568	beta R01_43	-3.085	0.217
16	5.5396413	NA	beta R01_47	2.263	0.381
Tongedächtnis: Personenparameter (k = 17)			Tongedächtnis: Item-Leichtigkeits-Parameter (k = 17)		
Raw Score	Estimate	Std. Error		Estimate	Std. Error
0	-4.15516826	NA	beta R02_01	2.186	0.331
1	-3.21610234	1.0835354	beta R02_02	-0.772	0.189
2	-2.34511434	0.8255880	beta R02_03	-0.772	0.189
3	-1.75398072	0.7233247	beta R02_04	1.509	0.275
4	-1.27175723	0.6704273	beta R02_05	-0.661	0.191
5	-0.84348855	0.6412282	beta R02_08	0.597	0.224
6	-0.44286813	0.6266800	beta R02_09	-0.193	0.199
7	-0.05326794	0.6234607	beta R02_10	0.881	0.238
8	0.33907427	0.6312061	beta R02_12	0.199	0.210
9	0.74914312	0.6519591	beta R02_13	-1.736	0.187
10	1.19757383	0.6908753	beta R02_14	0.822	0.235
11	1.71887801	0.7589581	beta R02_15	-0.549	0.192
12	2.38275634	0.8821907	beta R02_16	-3.190	0.216
13	3.38114711	1.1550908	beta R02_17	1.679	0.287
14	4.46016084	NA			
Rhythmusgedächtnis: Personenparameter (k = 16)			Rhythmusgedächtnis: Item-Leichtigkeits-Parameter (k = 16)		
Raw Score	Estimate	Std. Error		Estimate	Std. Error
0	-3.913397848	NA	beta R03_01	0.999	0.199
1	-3.036350359	1.0548829	beta R03_02	1.175	0.206
2	-2.227884417	0.7860472	beta R03_03	0.217	0.178
3	-1.701687939	0.6760110	beta R03_04	1.571	0.226
4	-1.287063043	0.6167291	beta R03_05	-0.045	0.174
5	-0.929748879	0.5815382	beta R03_06	0.875	0.195
6	-0.604712919	0.5604817	beta R03_07	-1.419	0.176
7	-0.297684764	0.5491387	beta R03_08	-0.076	0.173
8	0.001237493	0.5455126	beta R03_09	-0.777	0.169
9	0.300093539	0.5490195	beta R03_10	-0.867	0.170
10	0.606903494	0.5602174	beta R03_11	0.641	0.188
11	0.931519981	0.5810852	beta R03_12	-0.628	0.169
12	1.288131289	0.6160439	beta R03_13	0.567	0.186
13	1.701664572	0.6750756	beta R03_14	-1.324	0.174
14	2.226261898	0.7849174	beta R03_15	-0.013	0.174
15	3.032491328	1.0538211	beta R03_16	-0.897	0.170
16	3.907161715	NA			

Anmerkung. Für die Randwerte der Personenparameter sind keine Standardfehler eingetragen, da diese Werte im Berechnungsprogramm nur interpoliert werden können.

10 Überprüfung der Fragestellungen

Für die Analyse der Fragestellungen mit *SPSS Statistics 22.0* (IMB Corp., 2013) wurden die Personenparameter je Skala herangezogen sowie der Gesamtrohwert. Das Alpha-Niveau wurde bei zweiseitiger Fragestellung auf .05 festgesetzt.

10.1. Musikalische Expertise (Konstruktvalidität)

►a) Gibt es einen Zusammenhang zwischen der Leistung im Test musikalischer Fertigkeiten (TeMuK) und den Jahren im Instrumental- oder Gesangsunterricht sowohl für die Gesamtstichprobe als auch auf das Geschlecht bezogen?

Zur Überprüfung dieser Fragestellung wurden Korrelationen nach Pearson berechnet. Aufgrund der großen Stichprobengröße kann eine bivariate Normalverteilung der Messwerte angenommen werden (Bortz & Döring, 2006; Field, 2009). Die Ergebnisse für die Gesamtstichprobe und für die beiden Geschlechter getrennt sind *Tabelle 10.1* zu entnehmen. Alle Zusammenhänge waren positiv und signifikant.

Tabelle 10.1: Pearson-Korrelation zwischen der Leistung im TeMuK und den Jahren an Instrumental- oder Gesangsunterricht sowohl für die Gesamtstichprobe ($N = 174$) als auch nach dem Geschlecht getrennt ($n_{\text{weiblich}} = 93$, $n_{\text{männlich}} = 81$) bei $\alpha = .05$

Skalen	Gesamtstichprobe		Gruppen		
	<i>r</i>	<i>p</i>	Geschlecht	<i>r</i>	<i>p</i>
par_tonhöhe	.541	<.001*	männlich	.463	<.001*
			weiblich	.619	<.001*
par_lautstärke	.428	<.001*	männlich	.389	<.001*
			weiblich	.464	<.001*
par_zeitmaß	.354	<.001*	männlich	.422	<.001*
			weiblich	.284	.006*
par_tonged	.556	<.001*	männlich	.458	<.001*
			weiblich	.648	<.001*
par_rhythmusged	.316	. <.001*	männlich	.276	.031*
			weiblich	.358	<.001*
rohw_gesamt	.554	<.001*	männlich	.497	<.001*
			weiblich	.616	<.001*

Wird die Effektstärke nach Cohen (1988, zitiert nach Bortz & Döring, 2006) herangezogen, um die Relevanz der Zusammenhänge beurteilen zu können, so ist nach der Einteilung: kleiner Effekt: $r = 0.1$; mittlerer Effekt: $r = 0.3$; starker Effekt: $r = 0.5$ für alle Korrelationen ein mittlerer bis starker Effekt anzunehmen.

➤b) Gibt es einen Zusammenhang zwischen der Leistung im Test musikalischer Fertigkeiten (TeMuK) und dem selbst berichteten Musikerstatus sowohl für die Gesamtstichprobe als auch auf das Geschlecht bezogen?

Da die Selbsteinschätzung des Musikerstatus einem Rangordnungsschema entspricht, wurde für die Analyse der Fragestellung eine Spearman-Rangkorrelation durchgeführt. Die Gruppe der Musik-Studenten wurde den Profis zugeordnet, da für diese ein hohes Mindestmaß an musikalischen Fertigkeiten anzunehmen ist (Tabelle 10.2.)

Tabelle 10.2: Spearman-Rang-Korrelation zwischen der Leistung im TeMuK und dem selbstberichteten Musikerstatus sowohl für die Gesamtstichprobe ($N = 174$) als auch nach dem Geschlecht getrennt ($n_{\text{weiblich}} = 93$, $n_{\text{männlich}} = 81$) bei $\alpha = .05$

Skalen	Gesamtstichprobe		Gruppen		
	r	p	Geschlecht	r	p
par_tonhöhe	.615	<.001*	männlich	.576	<.001*
			weiblich	.646	<.001*
par_lautstärke	.416	<.001*	männlich	.446	<.001*
			weiblich	.390	<.001*
par_zeitmaß	.411	<.001*	männlich	.466	<.001*
			weiblich	.322	<.001*
par_tonged	.634	<.001*	männlich	.604	.002*
			weiblich	.662	<.001*
par_rhythmusged	.422	<.001*	männlich	.405	<.001*
			weiblich	.420	<.001*
rohw_gesamt	.631	<.001*	männlich	.586	<.001*
			weiblich	.625	<.001*

Alle Zusammenhänge wurden signifikant. Wird die Effektstärke nach Cohen (1988, zitiert nach Bortz & Döring, 2006) herangezogen, um die Relevanz der Zusammenhänge beurteilen zu können, so ist nach der Einteilung: kleiner Effekt: $r = 0.1$; mittlerer Effekt: $r = 0.3$; starker Effekt: $r = 0.5$ für alle Korrelationen ein mittlerer bis starker Effekt anzunehmen.

►c) Unterscheiden sich die Musikerstatus-Gruppen in aufsteigender Reihenfolge in ihren musikalischen Fertigkeiten (TeMuK)?

Für die Beantwortung der Fragestellung war die Berechnung von Varianzanalysen mit den eigentlich interessierenden Gruppenvergleichen in aufsteigender Reihenfolge geplant. Bezüglich der Skala Tonhöhe und der Rohwerte des Gesamttests war die Homoskedastizität jedoch nicht gegeben (Levene-Test: $p < .001^*$). Alternativ wurden hier Kruskal-Wallis-Tests durchgeführt, welche auch signifikant ausfielen (für Tonhöhe: $H(4) = 66.676$, $p < .001^*$; für Rohwerte Gesamttest: $F(4,169) = 74.924$, $p < .001^*$). Anschließend wurden auch hier in aufsteigender Reihenfolge die interessierenden Gruppenvergleiche berechnet (bei Heteroskedastizität Welch-Test). Für die restlichen Skalen war die Homoskedastizität mittels Levene-Test gegeben (Lautstärke: $p = .618$, Zeitmaß: $p = .574$, Tongedächtnis: $p = .122$; Rhythmusgedächtnis: $p = .345$) und alle folgenden Varianzanalysen wurden signifikant (Lautstärke: $F(4,169)=14.873$, $p < .001^*$, Zeitmaß: $F(4,169)=9.683$, $p < .001^*$, Tongedächtnis: $F(4,169) = 27.159$, $p < .001^*$; Rhythmusgedächtnis: $F(4,169) = 9.727$, $p < .001^*$).

Für einen Überblick der Mittelwerte je Gruppe und für alle Skalen siehe *Tabelle 10.3.* und *Abbildung 10.1 bis 10.6.* Da sowohl das Alter der Testpersonen in den Musikergruppen als auch das Geschlecht eine Rolle bei den Ergebnissen spielen könnte, werden im Folgenden die beiden Variablen im Bezug zum Musikerstatus in *Abbildung 10.7* und *10.8.* grafisch dargestellt und in *Tabelle 10.4* und *10.5* aufgelistet. Für weitere Berechnungen siehe *Kapitel 10.3.*

Anmerkung: An dieser Stelle ist kritisch anzumerken, dass ein gruppenweiser aufsteigender Vergleich mittels Kontraste eventuell überfrüht angewandt wurde. Alternativ wäre ein Student-Newman-Keuls-Test für die Identifizierung homogener Untergruppen anzudenken (vergl. Brosius, 1998).

Tabelle 10.3: Mittelwerte und Standardabweichung der gemessenen Fertigkeiten im TeMuK nach Gruppen getrennt

Musiker- Status- Gruppen	par_ tonhöhe		par_ lautstärke		par_ zeitmaß		par_ tonged		par_ rhythmusged		rw_ gesamt	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Nicht-Musiker	-.09	.23	-.20	.20	.25	.26	-.24	.26	-.38	.19	35.66	1.91
Hobby-Musiker	1.64	.20	1.27	.14	1.75	.17	1.32	.16	.71	.14	51.57	1.16
Semi-Profis	2.47	.44	.74	.3	1.99	.61	1.78	.49	.84	.46	53.21	2.82
Musik-Studenten	3.37	.13	1.94	.2	2.46	.22	3.12	.19	1.65	.26	62.29	1.05
Profis	3.06	.29	1.47	.31	1.73	.34	2.66	.27	.72	.42	56.82	1.86

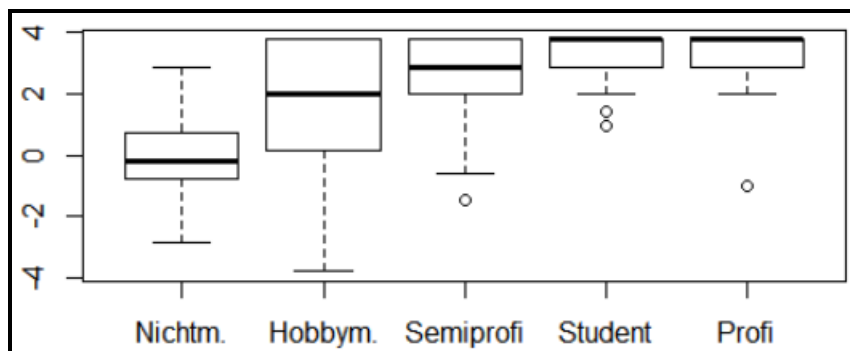


Abbildung 10.1: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Tonhöhe für die Gesamtstichprobe (N = 174)

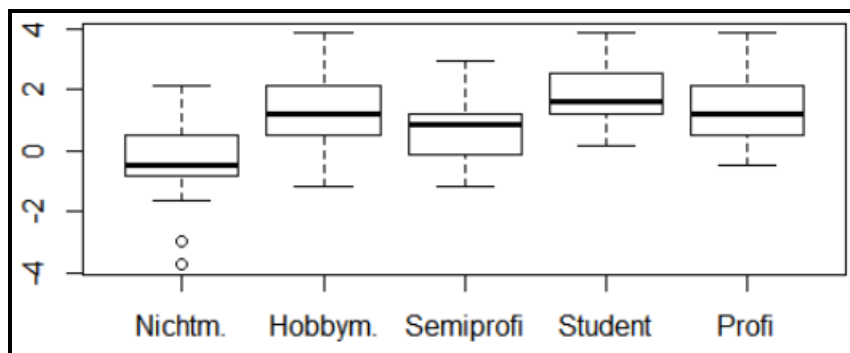


Abbildung 10.2: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Lautstärke für die Gesamtstichprobe (N = 174)

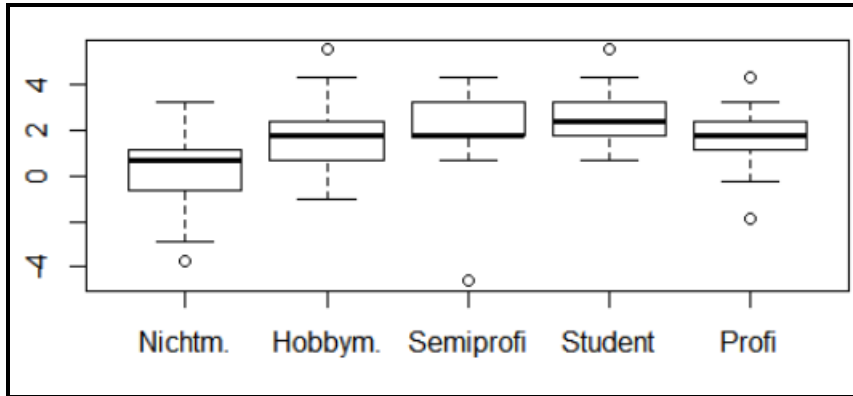


Abbildung 10.3: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Zeitmaß für die Gesamtstichprobe ($N = 174$)

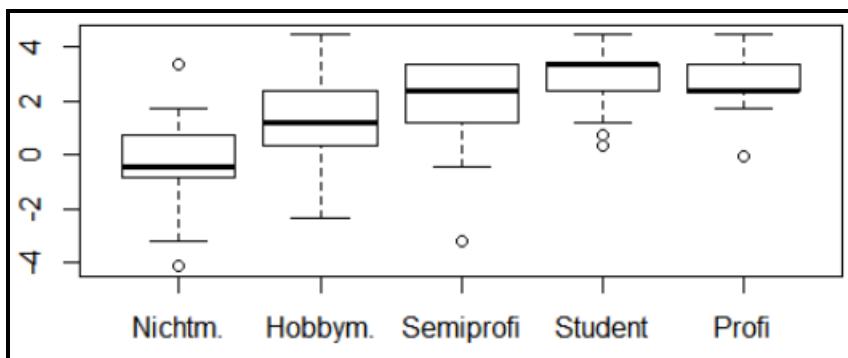


Abbildung 10.4: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Tongedächtnis für die Gesamtstichprobe ($N = 174$)

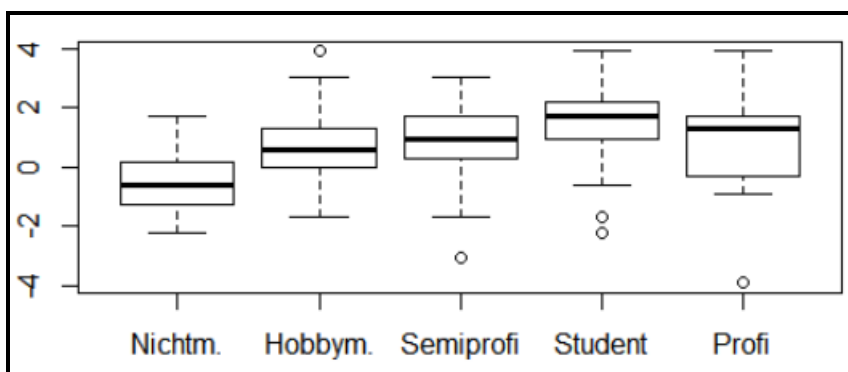


Abbildung 10.5: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Rhythmusgedächtnis für die Gesamtstichprobe ($N = 174$)

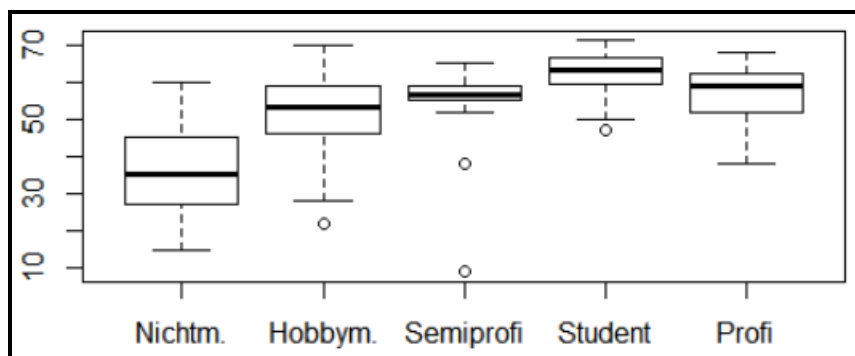


Abbildung 10.6: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen des über die Rohwerte aller Skalen ermittelten Summenscores für die Gesamtstichprobe ($N = 174$)

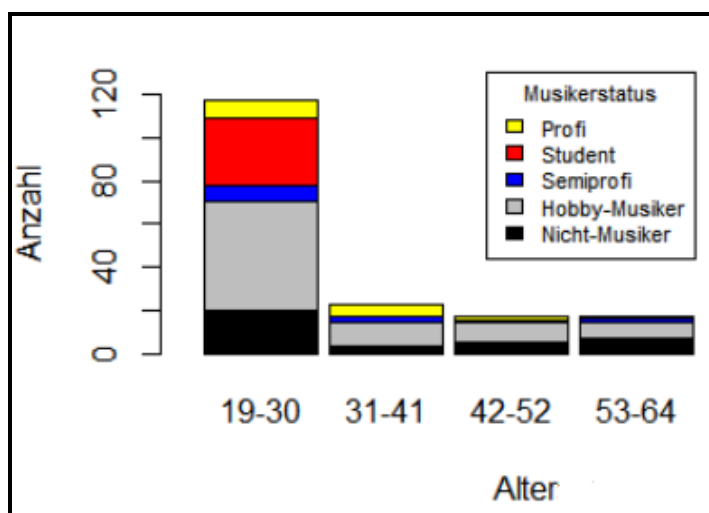


Abbildung 10.7: Altersverteilung der Musikergruppen

Tabelle 10.4: Altersverteilung nach Musikergruppen

Musikerstatus	Altersbereich			
	Häufigkeit der 19-30 jährigen	Häufigkeit der 31-41 jährigen	Häufigkeit der 42-52 jährigen	Häufigkeit der 53-64 jährigen
Nicht-Musiker	20	3	5	7
Hobby-Musiker	50	11	9	7
Semiprofi	8	3	1	2
Student	31	0	0	0
Profi	8	6	2	1
Gesamt ($N = 174$)	117	23	17	17

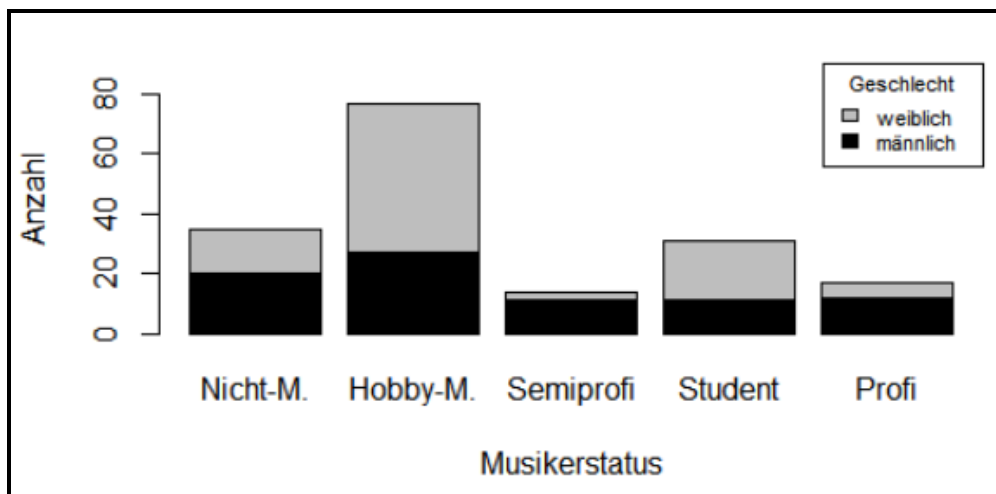


Abbildung 10.8: Altersverteilung der Musikergruppen

Tabelle 10.5: Geschlechterverteilung nach Musikergruppen

	Häufigkeit				
	Nicht-Musiker	Hobby-Musiker	Semiprofi	Student	Profi
Männlich	20	27	11	11	12
Weiblich	15	50	3	20	5
Gesamt	35	77	14	31	17

Die Ergebnisse zusammenfassend, kann folgendes ausgesagt werden:

Bei den folgenden Gruppenvergleichen zeigte sich für alle Skalen, dass Hobby-Musiker besser als Nicht-Musiker abschneiden und die Musik-Studenten von allen untersuchten Gruppen am besten performen. Weiters zeigte sich für die einzelnen Skalen im Speziellen (Tabelle 10.6 bis 10.11):

- (a) Die Semi-Profis unterscheiden sich hinsichtlich ihrer Leistung nicht von den Hobby-Musikern und sind nur in den Skalen Lautstärke und Tongedächtnis schlechter als die Musik-Studenten.
- (b) Die Leistung der Profis ist nur im Rhythmusgedächtnis schlechter als jene der Musik-Studenten. Dennoch zeigen die Profis nur in den Skalen Tonhöhe und Tongedächtnis bessere Leistungen als die Hobby-Musiker.
- (c) Bei Betrachtung der Rohwerte des Gesamttests zeigen sich auch Leistungsunterschiede zwischen Nicht-Musikern, Hobby-Musikern und Musik-

Studenten in aufsteigender Reihenfolge. Die Semi-Profis sind nicht besser als die Hobby-Musiker und erreichen einen niedrigeren Score als die Musik-Studenten. Die Profis heben sich zwar von den Hobby-Musikern ab, schneiden aber schlechter als die Musik-Studenten ab.

Tabelle 10.6: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Tonhöhe des TeMuK

Gruppe 1	n_1	Gruppe 2	n_2	Levene-Test		t-Test/ Welch-Test		
				F	p	t	df	p
Nicht-Musiker	35	Hobby-Musiker	77	9.100	.003*	-5.634	84.92	<.001*
Hobby-Musiker	77	Semi-Profis	14	1.932	.168	-1.601	89	.113
		Musik-Studenten	31	37.163	<.001*	-7.075	105.85	<.001*
		Profis	17	12.406	.001*	-4.044	34.63	<.001*
Semi-Profis	14	Musik-Studenten	31	7.913	.007*	-1.973	15.48	.067
Musik-Studenten	31	Profis	17	.857	.369	1.108	46	.274

Tabelle 10.7: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Lautstärke des TeMuK

Gruppe 1	n_1	Gruppe 2	n_2	t-Test		
				t	df	p
Nicht-Musiker	35	Hobby-Musiker	77	-5.954	169	<.001*
Hobby-Musiker	77	Semi-Profis	14	1.512		.132
		Musik-Studenten	31	-2.605		.010*
		Profis	17	-.602		.548
Semi-Profis	14	Musik-Studenten	31	-3.085		.002*
Musik-Studenten	31	Profis	17	1.302		.195

Tabelle 10.8: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Zeitmaß des TeMuK

Gruppe 1	n_1	Gruppe 2	n_2	t-Test		
				t	df	p
Nicht-Musiker	35	Hobby-Musiker	77	-4.807	169	<.001*
Hobby-Musiker	77	Semi-Profis	14	-.537		.592
		Musik-Studenten	31	-2.176		.031*
		Profis	17	.051		.959
Semi-Profis	14	Musik-Studenten	31	-.952		.342
Musik-Studenten	31	Profis	17	1.579		.116

Tabelle 10.9: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Tongedächtnis des TeMuK

Gruppe 1	n_1	Gruppe 2	n_2	t-Test		
				t	df	p
Nicht-Musiker	35	Hobby-Musiker	77	-5.452	169	<.001*
Hobby-Musiker	77	Semi-Profis	14	-1.128		.261
		Musik-Studenten	31	-6.037		<.001*
		Profis	17	-3.564		<.001*
Semi-Profis	14	Musik-Studenten	31	-2.970		.003*
Musik-Studenten	31	Profis	17	1.090		.277

Tabelle 10.10: Ergebnisse der Gruppenvergleich nach Musikerstatus im Bezug auf die Leistung in der Skala Rhythmusgedächtnis des TeMuK

Gruppe 1	n_1	Gruppe 2	n_2	t-Test		
				t	df	p
Nicht-Musiker	35	Hobby-Musiker	77	-4.010	169	<.001*
Hobby-Musiker	77	Semi-Profis	14	-.339		.735
		Musik-Studenten	31	-3.328		.001*
		Profis	17	-.042		.967
Semi-Profis	14	Musik-Studenten	31	-1.893		.060
Musik-Studenten	31	Profis	17	2.308		.022*

Tabelle 10.11: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung im Gesamttest nach Rohwerten des TeMuK ($\alpha = .05$)

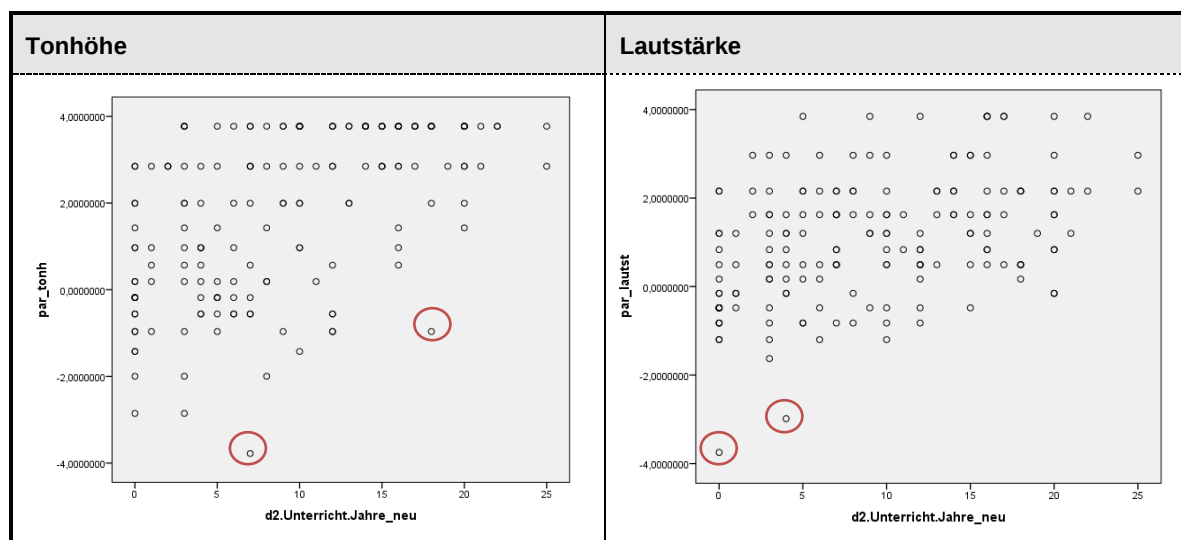
Gruppe 1	n_1	Gruppe 2	n_2	Levene-Test		t-Test/ Welch-Test		
				F	p	t	df	p
Nicht-Musiker	35	Hobby-Musiker	77	.870	.353	-7.387	110	<.001*
Hobby-Musiker	77	Semi-Profis	14	.067	.797	-.519	89	.605
		Musik-Studenten	31	10.054	.002*	-6.831	93.196	<.001*
		Profis	17	1.532	.220	-1.996	92	.049*
Semi-Profis	14	Musik-Studenten	31	3.721	.060	-3.046	43	.004*
Musik-Studenten	31	Profis	17	2.308	.136	2.765	46	.008*

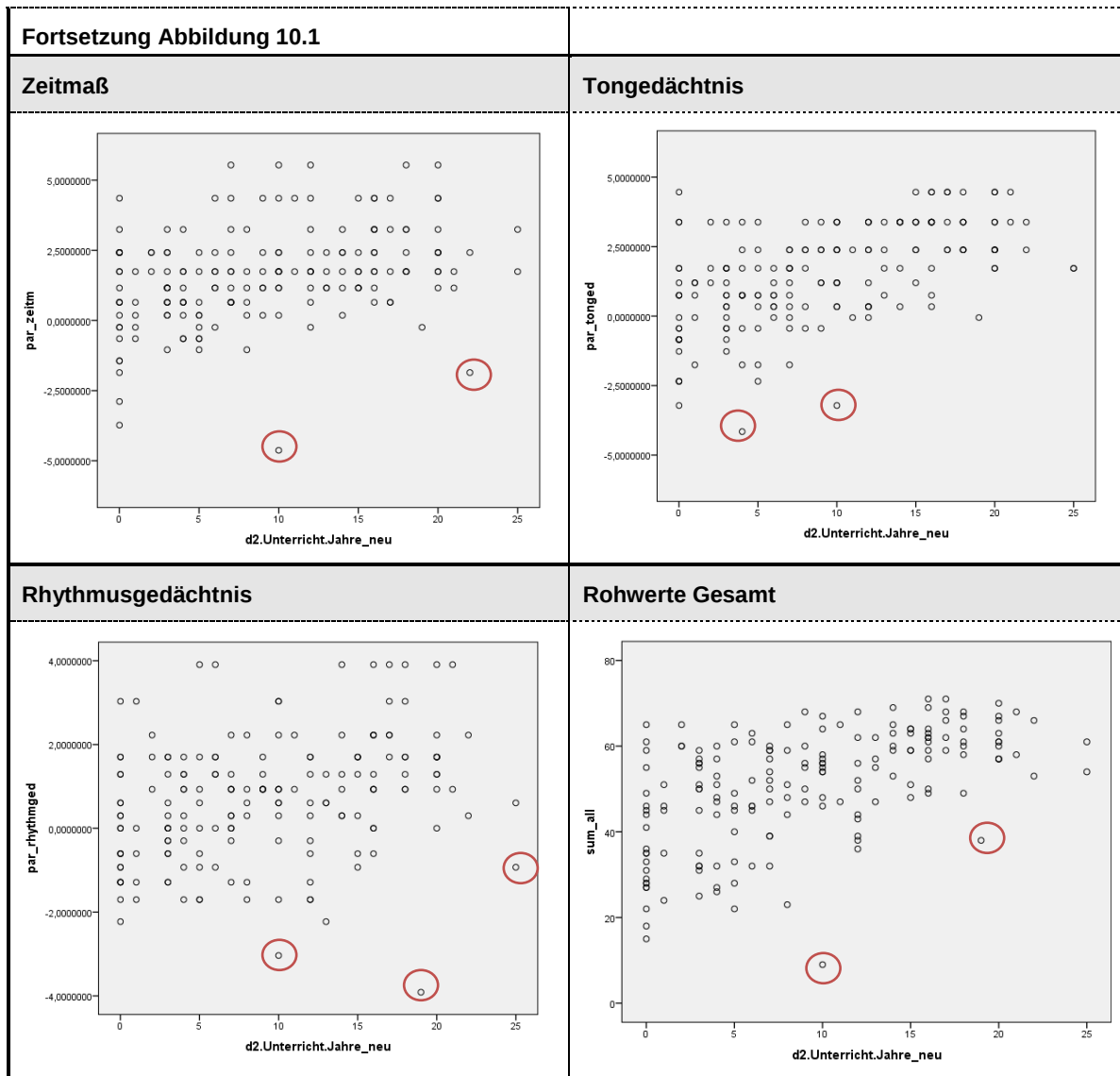
10.2. Sleeping musicians & musical sleepers

►Gibt es in der Stichprobe sleeping musicians oder musical sleepers nach Law und Zentner (2012)?

Zur Beurteilung werden die Streudiagramme für die Testleistung (Personenparameter) und der Unterrichtsjahre herangezogen (Abbildung 10.1) sowie die Boxplots aus den Abbildungen 10.1 bis 10.6. Es lassen sich sleeping musicians oder musical sleepers nach Law und Zentner (2012) augenscheinlich identifizieren, wobei die Frage nach der Ursache ungeklärt bleiben muss.

Abbildung 10.9: Streudiagramme für die Testleistung im TeMuK nach Personenparameter anhand der Anzahl der Jahre im Instrumental- oder Gesangsunterricht. Besonders auffallende Werte sind rot gekennzeichnet





10.3. Händigkeit, Geschlecht, Alter

► **a)** Gibt es einen Zusammenhang zwischen der Händigkeit und der Leistung im Test musikalischer Fertigkeiten (TeMuK) unabhängig von der musikalischen Expertise (Anzahl der Unterrichtsjahre)?

► **b)** Gibt es einen Zusammenhang zwischen dem Geschlecht und der Leistung im Test musikalischer Fertigkeiten (TeMuK) unabhängig von der musikalischen Expertise (Anzahl der Unterrichtsjahre)?

► **c)** Gibt es einen Zusammenhang zwischen dem Alter der Testpersonen und der Leistung im Test musikalischer Fertigkeiten (TeMuK) unabhängig von der musikalischen Expertise (Anzahl der Unterrichtsjahre)?

Zur Untersuchung des Zusammenhangs möglicher Einflussfaktoren wie Alter (jung = 0; alt = 1), Geschlecht (männl. = 0; weibl. = 1) und Händigkeit (li. = 0; re. = 1) wurde ein multiples Regressionsmodell erstellt. Für das Alter wurde zuvor auf Basis der Streudiagramme eine Einteilung in Jung (≤ 33) vs. Alt (≥ 34) vorgenommen. Als Kontrollvariable fungiert die Anzahl an Unterrichtsjahren. Um möglichen Suppressor-Effekten zu entgehen, wurde eine Rückwärts-Methode gewählt (Field, 2009). Im ersten Schritt wurden alle Variablen gleichzeitig in das Modell aufgenommen, in einem zweiten Schritt alle nicht signifikanten Variablen entfernt. Im Hinblick auf die Fragestellungen sind im Besonderen die partiellen Korrelationen interessant. Jedoch zeigen die Berechnungen für alle Skalen, dass lediglich die Anzahl der Unterrichtsjahre einen signifikanten Beitrag zur Vorhersage der Leistung leistet (Tabelle 10.10 bis 10.15). Das heißt weder über das Alter, das Geschlecht noch die Händigkeit kann ein Zusammenhang mit der Leistung im Test beobachtet werden.

Tabelle 10.12: Ergebnisse der Regressions-Analyse für die Skala Tonhöhe. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

	standardisierte Beta	$t(165)$	p	Pearson- Korrelation	partielle Korrelation
Konstante	.605	1.334	.184	-	-
Alter	.003	.051	.959	-.070	.004
Händigkeit	.012	.172	.864	-.086	.013
Geschlecht	-.108	-1.619	.107	-.096	-.125
Unterrichtsjahre	.527	7.821	>.001	.523	.520

Anmerkung. Multiples $R = .533$; $F(4,165) = 16.421$; $R^2 = .285$; $p < .001^*$; ohne Alter: $\Delta R^2 = .0$; $p = .959$; ohne Händigkeit: $\Delta R^2 = .0$; $p = .862$; ohne Geschlecht: $\Delta R^2 = -.011$; $p = .107$

Tabelle 10.13: Ergebnisse der Regressions-Analyse für die Skala Lautstärke. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

	standardisierte Beta	t(165)	p	Pearson-Korrelation	partielle Korrelation
Konstante	.531	1.525	.129	-	-
Alter	-.115	-1.604	.111	.171	-.124
Händigkeit	-.016	-.222	.825	-.081	-.017
Geschlecht	.024	-.329	.743	.032	.026
Unterrichtsjahre	.378	5.208	>.001	-.397	.376

Anmerkung. Multiples $R = .415$; $F(4,165) = 8.558$; $R^2 = .172$ $p < .001^*$; ohne Händigkeit: $\Delta R^2 = .0$; $p = .825$; ohne Geschlecht: $\Delta R^2 = .0$; $p = .762$; Alter: $\Delta R^2 = -.013$; $p = .105$

Tabelle 10.14: Ergebnisse der Regressions-Analyse für die Skala Zeitmaß. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

	standardisierte Beta	t(165)	p	Pearson-Korrelation	partielle Korrelation
Konstante	1.280	2.949	.004	-	-
Alter	.062	.843	.400	.015	.065
Händigkeit	-.050	-.675	.501	-.114	-.052
Geschlecht	-.118	-1.611	.108	-.120	-.124
Unterrichtsjahre	.325	4.345	>.001	.321	.321

Anmerkung. Multiples $R = .354$; $F(4,165) = 5.904$; $R^2 = .125$; $p < .001^*$, ohne Händigkeit: $\Delta R^2 = -.002$; $p = .501$; ohne Alter: $\Delta R^2 = -.004$; $p = .414$

Tabelle 10.15: Ergebnisse der Regressions-Analyse für die Skala Tongedächtnis. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

	standardisierte Beta	<i>t</i> (165)	<i>p</i>	Pearson-Korrelation	partielle Korrelation
Konstante	-.020	-.048	.962	-	-
Alter	.061	.935	.351	-.023	.073
Händigkeit	-.023	-.353	.724	-.096	-.027
Geschlecht	.101	1.559	.121	.108	.120
Unterrichtsjahre	.552	8.358	>.001	.548	.545

Anmerkung. Multiples $R = .561$; $F(4,165) = 18.906$; $R^2 = .314$; $p < .001^*$; ohne Händigkeit: $\Delta R^2 = -.001$; $p = .724$; ohne Alter: $\Delta R^2 = -.004$; $p = .356$; ohne Geschlecht: $\Delta R^2 = -.009$; $p = .133$

Tabelle 10.16: Ergebnisse der Regressions-Analyse für die Skala Rhythmusgedächtnis. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

	standardisierte Beta	<i>t</i> (165)	<i>p</i>	Pearson-Korrelation	partielle Korrelation
Konstante	.040	.100	.920	-	-
Alter	-.091	-1.225	.222	-.134	-.095
Händigkeit	.009	.123	.903	-.033	.010
Geschlecht	.071	.952	.343	.080	.074
Unterrichtsjahre	.286	3.783	>.001	.299	.283

Anmerkung. Multiples $R = .321$; $F(4,165) = 4.731$; $R^2 = .103$; $p < .001^*$; ohne Alter: $\Delta R^2 = .0$; $p = .903$; ohne Händigkeit: $\Delta R^2 = -.005$; $p = .329$; ohne Geschlecht: $\Delta R^2 = -.008$; $p = .214$

Tabelle 10.17: Ergebnisse der Regressions-Analyse für die Rohwerte des Gesamttests; Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

	standardisierte Beta	<i>t</i> (165)	<i>p</i>	Pearson-Korrelation	partielle Korrelation
Konstante	43.196	13.969	.000	-	-
Alter	-.019	-.282	.779	-.096	-.022
Händigkeit	-.029	-.424	.672	-.115	-.033
Geschlecht	-.010	-.153	.878	-.003	-.012
Unterrichtsjahre	.524	7.775	>.001	.531	.518

Anmerkung. Multiples $R = .533$; $F(4,165) = 16.346$; $p < .001^*$, $R^2 = .284$; ohne Geschlecht: $\Delta R^2 < .001$; $p = .878$; ohne Alter: $\Delta R^2 < .001$; $p = .781$; ohne Händigkeit: $\Delta R^2 = -.001$; $p = .647$

10.4. Selbsteinschätzung

►a) Gibt es einen Zusammenhang zwischen der Selbsteinschätzung und der Leistung im TeMuk sowohl für die Gesamtstichprobe als auch für die nach dem Geschlecht getrennte Stichprobe?

Zur Überprüfung der Fragestellung wurden Korrelationen nach Pearson berechnet. Hinsichtlich der großen Stichprobengröße kann eine bivariate Normalverteilung der Messwerte angenommen werden (Bortz & Döring, 2006; Field, 2009). Die Ergebnisse für die Gesamtstichprobe und für die nach dem Geschlecht geteilte Stichprobe sind aus *Tabelle 10.16* zu entnehmen. Alle Zusammenhänge wurden signifikant. Wird die Effektstärke nach Cohen (1988, zitiert nach Bortz & Döring, 2006) herangezogen, um die Relevanz der Zusammenhänge beurteilen zu können, so ist nach der Einteilung: kleiner Effekt: $r = 0.1$; mittlerer Effekt: $r = 0.3$; starker Effekt: $r = 0.5$ für alle Korrelationen ein mittlerer bis starker Effekt anzunehmen (sowohl für Männer als auch Frauen). Außerdem zeigt sich bei augenscheinlicher (!) Betrachtung der Höhe der Korrelationen, dass sich Männer und Frauen nicht wesentlich unterscheiden.

Tabelle 10.18: Pearson-Korrelation zwischen der Selbsteinschätzung der Begabung und der Leistung im TeMuK für die Gesamtstichprobe ($N = 174$) und nach dem Geschlecht getrennt ($n_{\text{weiblich}} = 93$, $n_{\text{männlich}} = 81$)

Skalen	Gesamtstichprobe		Gruppen		
	r	p	Geschlecht	r	p
par_tonhöhe	.527	<.001*	männlich	.548	<.001*
			weiblich	.508	<.001*
par_lautstärke	.433	<.001*	männlich	.438	<.001*
			weiblich	.437	<.001*
par_zeitmaß	.379	<.001*	männlich	.422	<.001*
			weiblich	.321	<.001*
par_tonged	.625	<.001*	männlich	.672	.002*
			weiblich	.611	<.001*
par_rhythmusged	.431	<.001*	männlich	.410	<.001*
			weiblich	.473	<.001*
rohw_gesamt	.630	<.001*	männlich	.636	<.001*
			weiblich	.630	<.001*

►b) Gibt es einen Zusammenhang zwischen der Selbsteinschätzung und dem Geschlecht unabhängig von der musikalischen Expertise?

Für die partielle Korrelation (Unterrichtsjahre als Kontrollvariable) zeigt sich kein signifikanter Zusammenhang zwischen Geschlechtes und Selbsteinschätzung unabhängig von der musikalischen Expertise ($r = -.122$; $p = .115$).

11 Diskussion & Ausblick

Ziel dieser Diplomarbeit war die Entwicklung und Validierung eines geeigneten Verfahrens zur Erfassung musikalischer Fertigkeiten (TeMuK) mit Hilfe des Rasch-Modells.

Auf Basis inhaltlicher Überlegungen, einer fundierten Literaturrecherche und zahlreicher Vorarbeiten nach Gittler (2009) wurde ein Online-Test zur Erfassung musikalischer Fertigkeiten (TeMuK) mit drei Wahrnehmungsskalen (Tonhöhe,

Lautstärke, Zeitmaß) und zwei Gedächtnisskalen (Tongedächtnis, Rhythmusgedächtnis) erstellt. 174 Testpersonen nahmen im Zeitraum von Jänner bis August 2014 an der Untersuchung teil. Darunter befanden sich Nicht-Musiker ($n = 35$), Hobby-Musiker ($n = 77$), Semiprofis ($n = 14$), Musikstudenten ($n = 31$) und Berufs- oder Profimusiker ($n = 17$). Zur Überprüfung der Rasch-Homogenität des TeMuK wurden je Skala sieben Modelltests nach Anderson (1973) durchgeführt (Teilungskriterien: Mittelwert, Median, Geschlecht, Alter, Bildungsstand, Musikerstatus, Jahre an Instrumental- bzw. Gesangsunterricht). Anschließend wurde zur weiteren Analyse Likelihood-Quotienten-Tests nach Martin Löff (1973) durchgeführt (Teilungskriterien: Median, Art der Aufgabenstellung, Itemreihenfolge).

Die Skalen Lautstärke, Zeitmaß und Rhythmusgedächtnis erwiesen sich als modellkonform. Für die Skala Tonhöhe konnte nach Ausschluss von drei Items (15, 45, 27) das Rasch-Modell angenommen werden. Die Modellgeltung ist jedoch kritisch zu betrachten. Bereits a priori galt die Vermutung, dass die wechselnden Tonhöhen der Items die Güte der Skala beeinflussen könnte. Bei der folgenden Item-Analyse zeigte sich, dass gerade diese Items einen Ton tiefer/höher als ihr vorhergehendes Item lagen und die Art der Veränderung in dieselbe Richtung gerichtet war. Abgesehen davon wurden diese Items bereits im Voraus als besonders schwierig eingestuft, da die Größe der Tonhöhenänderung für diese Aufgaben am geringsten war. Insbesondere bei den schwierigsten Items scheinen sich Personen mit wenig musikalischer Expertise von der Tonhöhe vorangegangener Items in ihrem Antwortverhalten beeinflussen zu lassen. Für die Skala Tonhöhe ist demnach ein Reihenfolge-Effekt durch die wechselnden Tonhöhen der Aufgaben anzunehmen, d.h. eine Verletzung der lokalen stochastischen Unabhängigkeit liegt vor.

Für die Skala Tongedächtnis war das Gütekriterium der Eindimensionalität nach Ausschluss der Items 6, 11, 7 erfüllt. Diese Items wurden bereits a priori als besonders schwierig bzw. kaum lösbar eingestuft, da der geänderte Ton dieselbe oder eine stark verwandte musikalische Stufenfunktion wie jener der Original-Melodie hat (für eine Erklärung siehe Amon, 2005). Anscheinend sind sich musikalisch erfahrene Personen dessen bewusst, ihnen fällt die Aufgabe relativ leichter (sie fühlen sich herausgefordert). Wenig erfahrene Personen bemerken Töne unterschiedlicher Tonhöhe aber gleicher Funktion relativ schwerer (sie lassen sich täuschen). Des Weiteren könnte auch ein gewisser Leittoncharakter der geänderten Töne dieser Items für unterschiedliche Lösungsstrategien verantwortlich sein. Dies ließe sich dahingehend interpretieren, dass musikalisch erfahrene Personen aus

musikpraktischen und –theoretischen Gründen auf Schlusswendungen etwas mehr sensibilisiert sind.

Laut den Modelltests nach Andersen (1973) und den Martin-Löf-Tests messen die Skalen zwei Skalen des Wahrnehmungs-Modul Lautstärke und Zeitmaß eine gemeinsame latente Fähigkeit sowie die beiden Skalen des zeitlichen Moduls Zeitmaß und Rhythmusgedächtnis zusammen. Abgesehen davon, dass die Frage der die Einordnung von Zeitmaß sowie der übrigen Skalen offen bleiben muss, sollten die Ergebnisse kritisch betrachtet werden. Stelzel (1979) kritisierte bereits, dass bei vorliegen zweier Rasch-homogener Subskalen in einem zusammengesetzten Test die Modelltests nicht sensitiv genug sind. Gittler (1999) geht diesem Problem bereits nach und zeigt auf wie es mittels spezieller Teilungsgruppenbildung im Andersen-Test (1973) und anderer trennscharfer Tests gelöst werden kann. Weitere Analysen des TeMuK sollten sich mit einer genügend großen Stichprobe daran orientieren. Vorerst wurde das Problem nur angedacht.

Insgesamt erwies sich der Test für Personen mit viel musikalischer Expertise als etwas zu leicht, wobei er sich für eine durchschnittliche Stichprobe sehr gut eignen könnte. Die Items erfassen den mittleren Fähigkeitsbereich sehr gut, die Randbereiche jedoch weniger differenziert. Person-Item-Maps geben Aufschluss über die Schwierigkeit der Items und somit Hinweise für die zukünftige Konstruktion von Items in spezifischen Schwierigkeitsbereichen.

Zur Überprüfung der Kriteriumsvalidität wurde der Zusammenhang musikalischer Expertise und der Testleistung untersucht. Für alle Skalen zeigte sich sowohl im Bezug auf die Unterrichtsjahre als auch den selbst berichteten Musiker-Status ein Zusammenhang mit mittleren oder starken Effekt. Die Untersuchung von Leistungsunterschieden ergab, dass Hobby-Musiker besser als Nicht-Musiker abschneiden und die Musik-Studenten von allen untersuchten Gruppen am besten performen. Die Profis oder Berufsmusiker zeigen nur in den Skalen Tonhöhe und Tongedächtnis bessere Leistungen als die Hobby-Musiker und schneiden beim Rhythmusgedächtnis schlechter ab als die Musik-Studenten. Anscheinend spielt die tonale Dimension im musikberuflichen Alltag eine besondere Rolle. Im Bezug auf die Rohwerte des Gesamttests heben sich die Profis zwar von den Hobby-Musikern ab, schneiden aber schlechter als die Musik-Studenten ab.

Anscheinend liegt ab einem gewissen Musikerstatus ein Plateau vor. Die Aktualität des laufenden Trainings der Studenten könnte für die guten Testergebnisse sprechen. Der Leistungsabfall ließe sich durch Gehörschäden von Berufsmusikern oder Alterseffekte

(Konfundierung) erklären. Bei der Regressionsanalyse zeigte sich kein Zusammenhang des Alters mit der Leistung, jedoch war die Stichprobe aufgrund Überrepräsentation junger Teilnehmer insbesondere in der Gruppe der Musikstudenten und der Hobby-Musiker nicht wirklich für diese Untersuchung geeignet. 117 Testpersonen der 174 befanden sich zum Erhebungszeitpunkt im Alter zwischen 19 und 30 Jahren. Eine speziell angelegte Querschnittsuntersuchung könnte im Bezug auf diese Fragestellung aufschlussreich sein, wobei Konfundierung auch oftmals ein Problem darstellt (Alter und Expertise). Eine Untersuchung des Zusammenhangs zwischen Geschlecht und musikalischer Testleistung mittels Regressionsanalyse ergab keinen Zusammenhang.

Trotz des Zusammenhangs musikalischer Fertigkeiten (im TeMuK) und musikalischer Expertise sind sowohl Personen mit ausgiebig musikalischer Erfahrung und einer dennoch geringen Personenfähigkeit (sleeping musicians), als auch einzelne Personen ohne bzw. mit geringer musikalischer Erfahrung und einer verhältnismäßig ausgeprägten Fertigkeit (musical sleepers) in der Stichprobe zu identifizieren. Jedoch stellt sich hier die Frage, ob vielleicht auch andere Gründe wie die Intensität der musikalischen Expertise, die Qualität des Unterrichts, die Tagesform, Störquellen im Raum, Motivation, Zeit, etc. hierfür verantwortlich sind.

Darüber hinaus kann über die Testgüte gesagt werden: Für den TeMuK gilt das Rasch-Modell. Die einzelnen Skalen erfassen die Facetten musikalischer Fertigkeiten eindimensional, die Beantwortung der Aufgaben ist voneinander unabhängig (für die Skala Tonhöhe kritisch zu betrachten), der Rohscore stellt ein verrechnungsfaires Maß dar, spezifisch objektive Vergleiche von Personen und Items sind möglich. Da es sich um einen computerbasierten Online-Test handelt und alle Personen mit derselben Aufforderung/Bitte rekrutiert wurden, kann ein hohes Maß an Durchführungsobjektivität angenommen werden. Auswertung der Ergebnisse wurde mittels freier Software *R.2.15.1.* (R Core Team, 2014) programmiert und erfüllen auch hier das Objektivitätskriterium. Im Bezug auf die Reliabilität kann das Kriterium der internen Konsistenz als erfüllt betrachtet werden. Sowohl LQT nach Anderson (1973), als auch nach Löff (1973) sprechen für die Homogenität der Items, sie messen dasselbe latente Konstrukt eindimensional. Die Leistung im TeMuK geht mit der Selbsteinschätzung der Teilnehmer einher, was für die Güte des Tests spricht. Männer und Frauen zeigen sich diesbezüglich gleichermaßen zuversichtlich (ein Zusammenhang zwischen Geschlecht und Selbsteinschätzung unabhängig von der musikalischen Expertise besteht nicht).

Kritisch zu betrachten wäre die Durchführung der Online-Testungen. Es stellt sich die Frage, inwiefern jede Testperson zu Hause wirklich ungestört war, alle annähernd gleiche technische Voraussetzungen hatten (PC, Qualität der Boxen, Lautstärke) und den Instruktionen folgten. Testabbrüche wurden nicht aufgezeichnet, welche bei überforderten Nicht-Musikern am ehesten zu vermuten sind. Bei den erfahrenen Musikern sind hingegen Deckeneffekte nicht auszuschließen. Zur Absicherung der Ergebnisse und Klärung offener Fragen sind weitere Studien notwendig.

Im Bezug auf die Händigkeit zeigte sich entgegen der Literatur kein Zusammenhang mit den erfassten musikalischen Fertigkeiten (Regressionsanalyse). Aufgrund der geringen Zahl an linkshändigen Teilnehmern ($n_1 = 19$) sind die Ergebnisse jedoch kritisch zu betrachten (Rechtshänder: $n_2 = 151$). Offen bleibt auch die Frage, ob hier kulturspezifische Fertigkeiten erfasst werden, oder der TeMuk auch in anderen Kulturen ein faires Testinstrument darstellt.

Der TeMuk kann als Versuch der Entwicklung des ersten computerbasierten Verfahrens zur Erfassung musikalischer Fertigkeiten auf Grundlage der probabilistischen Testtheorie (Spezialfall Rasch-Modell) angesehen werden. Eine erste Überprüfung der Testgüte kann als durchaus akzeptabel angenommen werden, wobei die Skala Tonhöhe im Bezug auf das Gütekriterium der lokalen stochastischen Unabhängigkeit kritisch zu betrachten ist. Weitere Studien sind zur Absicherung und Generalisierbarkeit der Ergebnisse dennoch notwendig.

Ausblick

Das Themengebiet der musikalischen Fertigkeiten stellte bislang in der Forschung einen vernachlässigten Intelligenzbereich dar. Es ist nur wenig über deren Struktur und Funktion bekannt (Gembris, 2009). Zukünftig besteht noch Forschungsbedarf im Bezug auf weitere Bereiche wie sprachliche Fertigkeiten, Emotionen oder Gedächtnisprozesse. Es gilt diese mit einem fairen Testinstrument näher zu untersuchen. Eine Ausweitung des TeMuK auf musikalische Wahrnehmungsbereiche wie Klangfarbe, Metrum oder musikanalytische Fertigkeiten zur Erfassung des Umgangs mit Motiven (kleinsten musikalischen Bausteinen) wäre ein Anliegen, um die musikalische Welt schrittweise besser verstehen zu können.

12 Verzeichnisse

12.1. Literaturverzeichnis

- Aggleton, J. P., Kentrige, R. W. & Good, J. M. M. (1994). Handedness and musical ability: A study of professional orchestral players, composers, and choir members. *Psychology of Music*, 22 (2), 148-156.
- Amon, R. (2005). *Lexikon: Der Harmonielehre*. Wien: Doblinger.
- Amuts, K., Jäncke, L., Mohlberg, H., Steinmetz H. & Zilles, K. (2000). Interhemispheric asymmetry of the human motor cortex related to handedness and gender. *Neuropsychologia*, 38, 304-312.
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38 (1), 123-140.
- Ayotte, J., Peretz, I. & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125, 238-251.
- Balaban, M. T., Anderson, L. M. & Wisniewski, A. B. (1998). Lateral asymmetries in infant melody perception. *Developmental Psychology*, 34 (1), 39-48.
- Baltes, P. B. (1990). Entwicklungspsychologie der Lebensspanne: Theoretische Leitsätze. *Psychologische Rundschau*, 41, 1-24.
- Bartlett, J.C. & Dowling, W.J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance*, 6, 501-515.
- Bear, M. F., Connors, B W. & Paradiso, M. A. (2007). *Neuroscience: Exploring the brain*. China: Lippincott Williams & Wilkins.
- Bentley, A. (1968). *Musikalische Begabung bei Kindern und ihre Meßbarkeit*. Frankfurt: Diesterweg.
- Best, C. T., Hoffman, H. & Glanville, B. B. (1982) Development of infant ear asymmetries for speech and music. *Perception & Psychophysics*, 31 (1), 75-85.
- Bever, T. G. & Chiarello, R.J. (1974). Cerebral dominance in musicians and nonmusicians. *Science*, 185, 537-539.
- Billroth, Th. (1895). *Wer ist musikalisch? Nachgelassene Schrift von Th. Billroth*, hrsg. von E. Hahnslick. Berlin: Gebrüder Paetel.

- Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics*, 53 (6), 585-600.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Bosnyak, D. J., Eaton, R. A. & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when nonmusicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex*, 14, 1088-1099.
- Brochard, R., Abecasis, D., Potter, D., Ragot, R. & Drake, C. (2003). The “Ticktock” of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, 14 (4), 362-366.
- Brosius, F. (1998). *SPSS 8*. Bonn: International Thomson Publishing.
- Catell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Colwell, R. & Abrahams, F. (1991): Edwin Gordon’s contribution: An appraisal. *The Quarterly Journal of music teaching and learning*, 2 (1-2), 19-36. (Reprint in *Visions of Research in Music Education*, 16 (2), 2010).
- Corrigall, K. A. & Trainor, L. J. (2009). Effects of musical training on key and harmony perception. *Annals of the New York Academy of Sciences*, 1169, 164-168. doi: 10.1111/j.1749-6632.2009.04769.x
- Cuddy, L. L., Balkwill, L. L., Peretz, I. & Holden, R. R. (2005). Musical difficulties are rare: A study of “Tone Deafness” among university students. *Annals of the New York Academy of Sciences*, 1060, 311-324. doi: 10.1196/annals.1360.026
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85 (4), 341-354.
- Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, 50 (4), 305-313.
- Dowling, W. J. & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 49, 524-531.
- Dowling, W. J. Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, 57 (2), 136-149.

- Field, A. (2009). *Discovering statistic using SPSS* (3rd rev. ed.). London: SAGE Publications Ltd.
- Fox, J. & Weisberg, S. (2011). *An {R} Companion to Applied Regression* (Second Edition). Thousand Oaks: Sage.
- Fujoka, T., Trainor, L. J., Ross, B., Kakigi, R. & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, 16 (6), 1010-1021.
- Furnham, A. (2001). Self-estimates of intelligence: Culture and gender differences in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 31, 1381-1405.
- Gaab, H., Keenan, J. P. & Schlaug, G. (2003). The effects of gender and the neural substrates of pitch memory. *Journal of Cognitive Neuroscience*, 15 (6), 810-820. doi: 10.1162/089892903322370735
- Gaser, C. & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, 23 (27), 9240-9245.
- Gembris, H. (1997). Historical phases in the definition of musicality. *Psychomusicology*, 16, 17-25.
- Gembris, H. (2005). Musikalische Begabung. In S. Helms, R. Schneider, & R. Weber (Hrsg.), *Lexikon der Musikpädagogik* (S. 31-33). Kassel: Bosse Verlag.
- Gembris, H. (2009). *Grundlagen musikalischer Begabung und Entwicklung* (3. unveränderte Aufl.). Augsburg: Wißner.
- Gittler, G. (1986). Inhaltliche Aspekte bei der Itemselektion nach dem Modell von Rasch. *Zeitschrift für experimentelle und angewandte Psychologie*, 33, 386-412.
- Gittler, G. (1999). Sind Raumvorstellung und Reasoning separierbare Fähigkeitsdimensionen? Dimensionalitätsanalyse zweier Rasch-skalierten Tests: 3DW und WMT. *Diagnostica*, 45 (2), 69-81.
- Gittler, G. (2009). TMB: *IRT-basierter Computertest zur Erfassung musikalischer Basisfertigkeiten* [Unveröffentlichte Forschungsversion]. Wien: Universität Wien.
- Gordon, E. E. (1965). *Musical Aptitude Profile*. Boston: Houghton Mifflin.
- Gordon, E. E. (1979). *Primary Measures of Music Audiation*. Chicago: G. I. A. Publications.

- Gordon, E. E. (1989a). *Advanced Measures of Music Audiation*. Chicago: G. I. A.
- Gordon, E. E. (1989b). *Audie*. Chicago: G. I. A. Publications.
- Grabowska, A., Gut, M., Binder, M., Forsberg, L., Rymarczyk, K., & Urbanik, A. (2012). Switching handedness: fMRI study of hand motor control in right-handers, left-handers and converted left-handers. *ACTA Neuropsychologiae Experimentalis*, 72, 439-451.
- Halpern, A. R. & Bartlett, J. C. (2002). Aging and memory for music: A review. *Psychomusicology*, 18, 10-27.
- Halpern, A. R., Bartlett, J. C. & Dowling, W. J. (1995). Aging and experience in the recognition of musical transpositions. *Psychology and Aging*, 10 (3), 325-342.
- Halpern, A. R., Bartlett, J. C. & Dowling, W. J. (1998). Perception of mode, rhythm, and contour in unfamiliar melodies: Effects of age and experience. *Music Perception*, 15 (4), 335-356.
- Halpern, A. R., Kwak, S. Y., Bartlett, J. C. & Dowling, W. J. (1996). The effects of aging and musical experience on the representation of tonal hierarchies. *Psychology and Aging*, 11, 235-246.
- Halpern, A. R. & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *The Quarterly Journal of Experimental Psychology*, 61 (9), 1371-1384.
- Hassler, M. & Gupta, D. (1993). Functional brain organization, handedness, and immune vulnerability in musicians and non-musicians. *Neuropsychologia*, 31 (7), 655-660.
- Hérbert, S. & Peretz I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition*, 25 (4), 518-533.
- Heygster, M. (2014). *Relative Solmisation: Grundlagen, Materialien, Unterrichtsverfahren*. Mainz: Schott.
- Highsmith, J. A. (1929). Selecting musical talent. *Journal of Applied Psychology*, 13 (5), 486-493.
- IBM Corp. (2013). *IBM SPSS Statistics for Windows* (Version 22.0). Armonk, New York: IBM Corp.
- Jäncke, L. (2009). *Macht Musik schlau? Neue Erkenntnisse aus den Neurowissenschaften und der kognitiven Psychologie*. Bern: Hogrefe AG.

- Jäncke, L., Schlaug, G. & Steinmetz, H. (1997). Hand skill asymmetry in professional musicians. *Brain and Cognition*, 34, 424-432.
- Johnson-Laird, P. N. (1991). Rhythm and meter: A theory at the computational level. *Psychomusicology*, 10, 88-106.
- Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, 41 (6), 621-634.
- Jones, M. R. & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96 (3), 459-491.
- Jones, M. R., Boltz, M. & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, 32 (3), 211-218.
- Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., Won, J. H., Worman, T., Yueh, B. & Rubinstein, J. (2009). Development and validation of the University of Washington Clinical Assessment of Music Perception Test. *Ear Hear*, 30 (4), 411-418. doi:10.1097/AUD.0b013e3181a61bc0.
- Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Föel, A., Ringelstein, E. B. & Henningsen, H. (2000). Handedness and hemispheric language dominance. *Brain*, 123, 2512-2518.
- Koelsch, S. (2011). Towards a neural basis of music perception- a review and updated model. *Frontiers in Psychology*, 2 (110). doi: 10.3389/fpsyg.2011.00110
- Koelsch, S., Gunter, T., Friederici, A. D. & Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of Cognitive Neuroscience*, 12 (3), 520-541.
- Koelsch, S., Maess, B., Grossmann, T. & Friederici, A. D. (2003). Electric responses reveal gender differences in music processing. *NeuroReport*, 14, 709-713.
- Koelsch, S. & Schröger, E. (2009). Neurowissenschaftliche Grundlagen der Musikwahrnehmung. In H. Bruhn, R. Kopiez & A.C. Lehmann (Hrsg.), *Musikpsychologie: Das neue Handbuch* (S. 393-412). Reinbek/Hamburg: Rowohlt.
- Koelsch, S., Schröger, E. & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport* 10, 1309-1313.
- Koelsch, S. & Siebel, W. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, 9 (12), 578-584.

- Koller, I., Alexandrowicz, R. & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis: Eine Einführung in eRem*. Wien: UTB.
- Kopiez, R., Galley, N. & Lee J. I. (2006). The advantage of a decreasing right-hand superiority: The influence of laterality on a selected musical skill (sight reading achievement). *Neuropsychologia*, 44, 1079-1087.
- Kries, J. v. (1926). *Wer ist musikalisch? Gedanken zur Psychologie der Tonkunst*. Berlin: Springer.
- Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2. Aufl.). Göttingen: Hogrefe.
- Law, L. N. C. & Zentner, M. (2012). Assessing musical abilities objectively: construction and validation of the Profile of Music Perception Skills. *PLoS ONE* 7 (12), e52508. doi:10.1371/journal.pone.0052508.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz PVU.
- Lynch, M. P., Eilers, R., E., Oller, D. K. & Urbano, R. C. (1990). Innateness, experience, and music. *Psychological Science*, 1 (4), 272-276.
- Mair, P., Hatzinger, R. & Maier, M. (2014). *eRm: Extended Rasch Modeling*. R package version 0.15-4.
- Martin-Löf, P. (1973). *Statistika Modeller: Anteckningar från seminarier Lasaret 1969-1970, utarbetade av Rolf Sunberg* [Statistical Models: Notes from seminars of the academic year 1969.1970, elaborated by Rolf Sunberg]. Obetydigt ändrat nytryck, oktober 1973. *Institutet för såringsmatematik och matematisk statistik vid Stockholms universitet*.
- McCarthy, D. (1930) A study of the Seashore measures of musical talent. *The Journal of Applied Psychology*, 14 (5), 437-455. doi: 10.1037/h0073360
- McLeish, J. (1968). The validity and reliability of Bentley's Measures of musical abilities. *British Journal of Educational Psychology*, 38 (2), 201.
- Merrett, D. L., Peretz, I. & Wilson, S. J. (2013). Moderating variables of music training-induced neuroplasticity: A review and discussion. *Frontiers in Psychology*, 4 (606). doi: 10.3389/fpsyg.2013.00606
- Miller, G. A. (1956). The magical seven, plus or minus two: Some Limits on Our capacity for Processing Information. *Psychological Review*, 63 (2), 81-97.

- Oerter, R. & Lehmann, A. C. (2009). Musikalische Begabung. In H. Bruhn, R. Kopiez & A.C. Lehmann (Hrsg.), *Musikpsychologie: Das neue Handbuch* (S. 88-104). Reinbek/Hamburg: Rowohlt.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles*, 48 (5-6), 265-276.
- Parncutt, R. (2009). Prenatal and infant conditioning, the mother schema, and the origins of music and religion. *Musicae Scientiae*, 13, 119-150. doi: 10.1177/1029864909013002071
- Peretz, I. (1990). Processing of local and global musical information in unilateral brain-damaged patients. *Brain*, 113 (4), 1185-1205).
- Peretz, I. & Babaii, M. (1992). The role of contour and interval in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, 30 (3), 277-292.
- Peretz, I., Champod, A. S. & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals New York Academy of Sciences*, 999, 58-75.
- Peretz, I. & Gosselin, N. (2008). On-line identification of congenital amusia. *Music Perception*, 25 (4), 331-343.
- Peretz, I. & Kolinsky, R. (1993). Boundaries of seperability between melody and rhythm in music discrimination: A neuropsychological perspective. *Quarterly Journal of Experimental Psychology*, 46, 301-325.
- Peretz, I. & Zatorre, R. J. (2005): Brain organization for music processing. *Annual Review of Psychology*, 56, 89-114.
- Piro, J. & Ortiz, C. (2010). No association between music ability and hand preference in children. *Journal of Motor Behavior*, 42 (5), 269-275.
- Prichard, E., Propper, R. E. & Christman, S. D. (2013). Degree of handedness, but not direction, is a systematic predictor of cognitive performance. *Frontiers in Psychology*, 4 (9). doi: 10.3389/fpsyg.2013.00009
- Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L. & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 29, 191-212.

- Radvanscy, G. A. & Potter, J. K. (2000). Source cuing: Memory for melodies. *Memory & Cognition*, 28 (5), 693-699.
- Rammstedt, B. & Rammsayer, T. H. (2000). Sex differences in self-estimates of different aspects of intelligence. *Personality and Individual Differences*, 29, 869-880.
- R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ruthsatz, J., Detterman, D., Griscom, W. S. & Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. *Intelligence*, 63, 330-338.
- Saunders, T.C. (1991). The stages of music audiation: A survey of research. *The Quarterly*, 2 (1-2), 131-137. (Reprinted with permission in *Visions of Research in Music Education*, 16 (2), 2010).
- Schaub, S. (1984). *Methodenbeiträge zur Erforschung des Musik-Lernens. Die Erfassung musikalischen Erlebens und musikalischer Einstellungen und deren Bedeutung für das Konzept der Musikalität*. Mainz: Schott.
- Schellenberg, E., G. & Moreno, S. (2009). Music lessons, pitch processing, and g. *Psychology of Music*, 38 (2), 209-221.
- Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaption. *Annals of the New York Academy of Sciences*, 930, 281-299.
- Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F. & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, 33 (8), 1047-1055.
- Schlaug, G., Jäncke, L., Huang, Y. & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, 267 (5198), 699-701.
- Schlaug, G., Norton, A., Overy, K. & Winner, A. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, 1060, 219-230. doi: 10.1196/annals.1360.015.
- Schneider, P., Scherg, M., Dosch, H.G., Specht, H.-J., Gutschalk, A. & Rupp, A. (2002). Morphology of Heschls gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, 5 (7), 688-694.

- Schneider, P., Sluming, V., Roberts, N., Bleeck, S. & Rupp, A. (2005). Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Annals of the New York Academy of Sciences*, 1060, 387-394.
- Schulze, K., Dowling, W. J. & Tillmann, B. (2012). Working memory for tonal and for atonal sequences during a forward and a backward recognition task. *Music Perception*, 29 (3), 255-267. doi:10.1525/MP.2012.29.3.255
- Schuppert, M. (2009). Amusie: Störungen der Musikverarbeitung. In H. Bruhn, R. Kopiez & A.C. Lehmann (Hrsg.), *Musikpsychologie: Das neue Handbuch* (S. 613- 629). Reinbek/Hamburg: Rowohlt.
- Schuppert, M., Münte, T. F., Wieringa, B. M. & Altenmüller, E. (2000). Receptive amusia: Evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain*, 123 (3), 546-559.
- Seashore, C. E. (1919). *Seashore measurement of musical talent*. New York, revidierte Fassungen 1939, 1956, 1960; deutsche Ausgabe hg. von H. Fischer & Ch. Butch (1966): Bern.
- Seashore, C. E. (1938). *Psychology of music*. New York: McGraw-Hill Book Company, Inc.
- Shuter-Dyson, R. (1980). Psychometrische und experimentelle Studien zur musikalischen Begabung. In K. E. Behne (Hrsg.), *Einzeluntersuchungen* (Musikpädagogische Forschung, Bd. 1) (S. 50-59). Laaber: Laaber.
- Silverman, M. J. (2010). The effect of pitch, rhythm, and familiarity on working memory and anxiety as measured by digit recall performance. *Journal of Music Therapy*, 47 (1), 70-83.
- Sloboda, J. A. (1993). Musical ability. In G. R. Bock & K. Ackrill (Eds.), *The origins and development of high ability* (pp.106-118). Chichester: Wiley.
- Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A. & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage*, 17 (3), 1613-1622.
- Spearman, C. (1904). 'General intelligence', objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spitzer, M. (2009). *Musik im Kopf: Hören, Musizieren, Verstehen und Erleben im neuronalen Netzwerk*. Stuttgart: Schattauer.

- Stefani, G. (1987). A theory of musical competence. *Semiotica*, 66 (1), 7-22.
- Stelzel, I. (1979) Ist der Modelltest des Rasch-Modells geeignet Homogenitätshypothesen zu prüfen? Ein Bericht über Simulationsstudien mit inhomogenen Daten. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 26, 625-672.
- Stemmler, G., Hagemann, D., Amelang, M. & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung* (7. Aufl.). Kohlhammer: Stuttgart.
- Stewart, L., Kriegstein, K., Warren, J. D. & Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain*, 129, 2533-2553.
- Tervaniemi, M., Castaneda, A., Knoll, M. & Uther, M. (2006). Sound processing in amateur musicians and nonmusicians: Event-related potential and behavioral indices. *Cognitive Neuroscience and Neuropsychology*, 17 (11), 1225-1228.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A. & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs. nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161, 1-10. doi: 10.1007/s00221-004-2044-5
- Thunks, T. W., Bowers, D. R. & Eagle, C. T. (1993). The effect of stimulus tempo on melodic error detection. *Psychomusicology*, 12, 41-51.
- Tillmann, B., Bharucha, J. J. & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, 107 (4), 885-913.
- Trainor, L. J., Desjardins, R. N. & Rockel, C. (1999). A comparison of contour and interval processing in musicians and nonmusicians Using event-related potentials. *Australian Journal of Psychology*, 51 (3), 147-153.
- Trainor, L. J. & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18 (2), 394-402.
- Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences*, 930, 1-16.
- Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, 6 (7), 669-673.

- Vanecek, E., Aßlaber, M., Längle, P. & Preusche, I. (2004). *Der Wiener Test für Musikalität (WTM). Computerbasierter Musikbegabungstest für Kinder im Vor- und Volksschulalter*. ARC Seibersdorf: Austrian Research Centers.
- Verghese, J., Lipton, R. B., Katz, M. J., Hall, C. B., Derby, C. A., Kuslansky, G., et al. (2003). Leisure activities and the risk of dementia in the elderly. *The New England Journals of Medicine*, 348 (25), 2508-2516.
- Wing, H. D. (1961). *Standardised tests of musical intelligence*. The mere: England: National Foundation for Educational Research.

12.2. Tabellenverzeichnis

Tabelle 4.1: Studien zur Untersuchung der hemisphärischen Verarbeitung von musikalischer Information

Tabelle 8.1: Musikalischen Expertise bzw. Musikerstatus der Gesamtstichprobe mit Häufigkeiten und Prozentanteilen; Durchschnittswerte der Unterrichtsjahre und der Selbsteinschätzung der Begabung je Gruppe

Tabelle 9.1: Ergebnisse des 1. Durchgang der Modelltests (Andersen, 1973) für die Subskala Tonhöhe, $N = 174$, $k = 16$, $\alpha = .01$

Tabelle 9.2: Ergebnisse des letzten Durchgang des Modelltests (Andersen, 1973) für die Subskala Tonhöhe nach Ausschluss der Items 15, 45, 27 bei $N = 174$, $k = 13$, $\alpha = .01$

Tabelle 9.3: Ergebnisse des Modelltests (Andersen, 1973) für die Subskala Lautstärke bei $N = 174$, $k = 15$, $\alpha = .01$

Tabelle 9.4: Ergebnisse des Modelltests (Andersen, 1973) für die Subskala Zeitmaß bei $N = 174$, $k = 16$, $\alpha = .01$

Tabelle 9.5: Ergebnisse des 1. Durchgang der Modelltests (Andersen, 1973) für die Subskala Tongedächtnis bei $N = 174$, $k = 17$, $\alpha = .01$

Tabelle 9.6: Ergebnisse des letzten Durchgang des Modelltests (Andersen, 1973) für die Subskala Tongedächtnis nach Ausschluss der Items 6, 11, 7 bei $N = 174$, $k = 14$, $\alpha = .01$

Tabelle 9.7: Ergebnisse des der Modelltests (Andersen, 1973) für die Subskala Rhythmusgedächtnis bei $N = 174$, $k = 16$, $\alpha = .01$

Tabelle 9.8: Ergebnisse des der Modelltests (Andersen, 1973) für die Subskalen Zeitmaß und Lautstärke zusammen bei $N = 174$, $k = 31$, $\alpha = .01$

Tabelle 9.9: Ergebnisse des der Modelltests (Andersen, 1973) für die Subskalen Zeitmaß und Rhythmusgedächtnis zusammen bei $N = 174$, $k = 32$, $\alpha = .01$

Tabelle 9.10: Ergebnisse des Test nach Martin-Löf (1973) für die Subskala Tonhöhe nach Ausschluss der Items 15, 45 und 27 ($k = 13$)

Tabelle 9.11: Ergebnisse des Martin-Löf-Tests (1973) für die Subskala Lautstärke ($k = 15$)

Tabelle 9.12: Ergebnisse des Test nach Martin-Löf (1973) für die Subskala Zeitmaß ($k = 16$)

Tabelle 9.13: Ergebnisse des Test nach Martin-Löf (1973) für die Subskala Tongedächtnis nach Ausschluss der Items 6, 11 und 7 ($k = 14$)

Tabelle 9.14: Ergebnisse des Test nach Martin-Löf (1973) für die Subskala Rhythmusgedächtnis ($k = 16$)

Tabelle 9.15: Ergebnisse des Test nach Martin-Löf (1973) für die Subskalen Lautstärke und Zeitmaß zusammen ($k = 31$)

Tabelle 9.16: Ergebnisse des Test nach Martin-Löf (1973) für die Subskalen Zeitmaß und Rhythmusgedächtnis zusammen ($k = 32$)

Tabelle 9. 17: Häufigkeiten der Parameterwerte in der Gesamtstichprobe ($N = 174$)

Tabelle 9.18: Personenparameter und Item-Leichtigkeitsparameter aller Skalen sowie deren Standardfehler für die untersuchte Stichprobe $N = 174$

Tabelle 10.1: Produkt-Moment-Korrelation nach Pearson zwischen der Leistung im TeMuK und den Jahren an Instrumental- oder Gesangsunterricht sowohl für die Gesamtstichprobe ($N = 174$) als auch nach dem Geschlecht getrennt ($n_{\text{weiblich}} = 93$, $n_{\text{männlich}} = 81$) bei $\alpha = .05$

Tabelle 10.2: Spearman-Rang-Korrelation zwischen der Leistung im TeMuK und dem selbstberichteten Musikerstatus sowohl für die Gesamtstichprobe ($N = 174$) als auch nach dem Geschlecht getrennt ($n_{\text{weiblich}} = 93$, $n_{\text{männlich}} = 81$) bei $\alpha = .05$

Tabelle 10.3: Mittelwerte und Standardabweichung der gemessenen Fertigkeiten im TeMuK nach Gruppen getrennt

Tabelle 10.4: Altersverteilung nach Musikergruppen

Tabelle 10.5: Geschlechterverteilung nach Musikergruppen

Tabelle 10.6: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Tonhöhe des TeMuK

Tabelle 10.7: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Lautstärke des TeMuK.

Tabelle 10.8: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Zeitmaß des TeMuK

Tabelle 10.9: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Tongedächtnis des TeMuK

Tabelle 10.10: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung in der Skala Rhythmusgedächtnis des TeMuK

Tabelle 10.11: Ergebnisse der Gruppenvergleiche nach Musikerstatus im Bezug auf die Leistung im Gesamttest nach Rohwerten des TeMuK

Tabelle 10.12: Ergebnisse der Regressions-Analyse für die Skala Tonhöhe. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

Tabelle 10.13: Ergebnisse der Regressions-Analyse für die Skala Lautstärke. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

Tabelle 10.14: Ergebnisse der Regressions-Analyse für die Skala Zeitmaß. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

Tabelle 10.15: Ergebnisse der Regressions-Analyse für die Skala Tongedächtnis. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

Tabelle 10.16: Ergebnisse der Regressions-Analyse für die Skala Rhythmusgedächtnis. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

Tabelle 10.17: Ergebnisse der Regressions-Analyse für die Rohwerte des Gesamttest. Koeffizienten des ersten Modells und Pearson- sowie partielle Korrelationen der Variablen Geschlecht, Alter und Händigkeit

Tabelle 10.18: Pearson-Korrelation für die Selbsteinschätzung der Begabung und der Leistung im TeMuK für die Gesamtstichprobe ($N = 174$) und nach dem Geschlecht getrennt ($n_{\text{weiblich}} = 93$, $n_{\text{männlich}} = 81$)

12.3. Abbildungsverzeichnis

Abbildung 1.1: Typen und Stadien der Audiation (Gordon 1989, zitiert nach Saunders, 1991, S. 132)

Abbildung 8.1: Altersverteilung der Gesamtstichprobe ($N = 174$)

Abbildung 8.2: Ausbildungsstand der Gesamtstichprobe ($N = 174$)

Abbildung 8.3: Musikerstatus und Jahre an Instrumental- bzw. Gesangunterricht je Gruppe dargestellt an den absoluten Häufigkeiten in der Stichprobe ($N = 174$)

Abbildung 8.4: Musikerstatus und Einschätzung der Begabung je Gruppe dargestellt an den absoluten Häufigkeiten in der Stichprobe ($N = 174$)

Abbildung 8.5: Fragebogen zur musikalischen Expertise

Abbildung 8.6: Grafik zur Aufgabenerklärung für die Skala Zeitmaß

Abbildung 9.1: Person-Item-Map für die Subskala Tonhöhe

Abbildung 9.2: Person-Item-Map für die Subskala Lautstärke

Abbildung 9.3: Person-Item-Map für die Subskala Zeitmaß

Abbildung 9.4: Person-Item-Map für die Subskala Tongedächtnis

Abbildung 9.5: Person-Item-Map für die Subskala Rhythmusgedächtnis

Abbildung 9.6: Verteilung der Personenparameter der einzelnen Skalen sowie des Gesamt-Rohwerts für die Gruppe der Nicht-Musiker ($n = 35$) und Hobby-Musiker ($n = 77$)

Abbildung 10.1: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Tonhöhe für die Gesamtstichprobe ($N = 174$)

Abbildung 10.2: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Lautstärke für die Gesamtstichprobe ($N = 174$)

Abbildung 10.3: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Zeitmaß für die Gesamtstichprobe ($N = 174$)

Abbildung 10.4: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Tongedächtnis für die Gesamtstichprobe ($N = 174$)

Abbildung 10.5: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen der Personenparameter der Skala Rhythmusgedächtnis für die Gesamtstichprobe ($N = 174$)

Abbildung 10.6: Boxplot mit Mittelwerten, Quartilen und Standardabweichungen des über die Rohwerte aller Skalen ermittelten Summenscores für die Gesamtstichprobe ($N = 174$)

Abbildung 10.7: Altersverteilung der Musikergruppen

Abbildung 10.8: Altersverteilung der Musikergruppen

Abbildung 10.9: Streudiagramme für die Testleistung im TeMuK nach Personenparameter anhand der Anzahl der Jahre im Instrumental- oder Gesangsunterricht. Auffallende Werte sind rot gekennzeichnet

13 Anhang

13.1. Abstract - Deutsch

Das Ziel dieser Arbeit war die Entwicklung eines Forschungs- und Messinstruments zur Erfassung musikalischer Fertigkeiten (TeMuK) auf Basis zahlreicher Vorarbeiten (Gittler, 2009). Bisherige Musikalitätstests wurden auf Basis der klassischen Testtheorie konstruiert und beinhalten zahlreiche Mängel (Testtheorie, Qualität). Um dem entgegen zu wirken, wurde ein computerbasierter Musikalitäts-Test (TeMuK) auf Grundlage der probabilistischen Testtheorie (Rasch-Modell) entwickelt, welcher drei musikalische Wahrnehmungsbereiche (Tonhöhe, Lautstärke, Zeitmaß) und zwei Aspekte des Musikgedächtnisses (Ton- und Rhythmusgedächtnis) erfasst. 174 Testpersonen (81 Männer, 93 Frauen) wurden mit einer Online-Version im Einzelsetting getestet. Anschließend wurde das Messinstrument einer Rasch-Modellprüfung unterzogen. Nach Ausschluss einiger Items konnte für alle Skalen separat das Rasch-Modell angenommen werden. Auch für die Skalenkombination Rhythmusgedächtnis und Zeitmaß sowie für die Kombination Tondauer und Lautstärke gilt das Modell. Die Wahrnehmungs- und Gedächtnisskalen zusammen erfassen jedoch keine übergeordnete musikalische Fertigkeit. Im Bezug auf die Kriteriumsvalidität zeigte sich ein Zusammenhang zwischen Testleistung und musikalischer Expertise (Unterrichtsjahre, Musikerstatus). Des Weiteren ließ sich ein Zusammenhang zwischen der Testleistung und der Selbsteinschätzung der Begabung beobachten. Die Untersuchung von Leistungsunterschieden ergab, dass Hobby-Musiker besser als Nicht-Musiker abschneiden und die Musik-Studenten von allen untersuchten Gruppen am besten performten. Ein Zusammenhang der Testleistung mit Händigkeit, Alter oder Geschlecht konnte nicht gefunden werden.

Schlüsselwörter: Alter, Geschlecht, Händigkeit, Musiktest, musikalische Expertise, musikalische Fertigkeiten, Rasch-Modell, Testkonstruktion

13.2. Abstract - English

Existing tests of musical ability have primarily drawn on Classical Test Theory (CTT) and thus lead to a number of test theoretical problems. The aim of this study was to overcome these limitations by constructing a test of musical aptitude (Test zur Erfassung musikalischer Fertigkeiten; TeMuK) on the basis of probabilistic test theory and preliminary studies (Gittler, 2009). Three domains of perception (pitch, loudness and tempo) and two aspects of memory (melody and rhythm) were evaluated. The online version of the test was administered to 174 subjects. The instrument showed good fit to the Rasch model. After the exclusion of a number of items, Rasch conformity could be adopted for each separate scale. Equally, the scales of loudness and tempo taken together as well as the scales of rhythm and tempo taken together demonstrated good Rasch model fit. The scales of perception and memory, however, do not measure a related higher musical ability. Criterion validity was examined by calculating the correlation between musical training and musicianship status. Furthermore, the performance went along with the self-assessment of musical aptitude. A comparison of the groups indicated that hobby musicians performed better than non-musicians and musical students outperformed all groups taking part in the study. Overall, performance was not shown to be related to handedness, age or sex.

Key words: age, handedness, musical aptitude, music test, musical experience, Rasch model, sex, test construction

13.3. Lebenslauf

Angaben zur Person

Name: Karoline Lang, geb. am 14.03.1990, Hainburg/Donau

Ausbildung:

2004-2009 ORG für Studierende der Musik, Wien, Neustiftgasse
Juni 2009 Abitur/Matura mit ausgezeichnetem Erfolg
seit Herbst 2009 **Universität Wien: Studium Psychologie**
Schwerpunkt Klinische- und Gesundheitspsychologie
Universität für Musik u. Darstellende Kunst Wien:
Studium Instrumental- u. Gesangspädagogik (Querflöte)
Schwerpunkt: Musikphysiologie in Theorie und Praxis (u.a. Yoga, Pilates, Atem-, Haltungs- und Bewegungsschulung u.a. nach F.M. Alexander)

Psychologische Praktika:

2010 **SVB: Gesundheitsmaßnahme f. Kinder u. Jugendliche (Behandlung: Legasthenie, Logopädie)**
Kinderbetreuung mit erlebnispädagogischem Schwerpunkt, Einsatz therapeutischer Elemente u. Stärkung zwischenmenschlichen Verhaltens

2011 **SVB: Gesundheitsmaßnahme „Fit ins Leben“ für übergewichtige Kinder**
Mitarbeit u. Teilnahme bei d. Gruppentherapie Fokus: Selbstbild, Selbstwert, Hänselei, Frust, Ärger, Selbstreflexion; Leitende Psychologin: Dr. E. Hefler

2011 **Kompositionsprojekt e-compose Austria (Uni: mdw)**
Erprobung einer digitalen internetbasierten Lernplattform an österr. Schulen; Projektmitarbeit als Musikpädagogin, Workshop-Begleitung

2012 **LeFoP: Lehr- und Forschungspraxis Uni Wien**
Forschungspraktikum: Anwendung klinisch-psychologischer Testverfahren, Behandlungs- u. Interventionsmethoden, VR-Bedienung, Forschungs-Mitarbeit

Berufserfahrung:

seit 2000 **musikalische Auftritte**
u. a. Grafenegg, Universitätskapelle der BOKU etc.

seit 2010 **geringfügige Beschäftigung im landwirtschaftlichen Betrieb**
Land- u. Feldarbeit, administrative Tätigkeiten

Hiermit bestätige ich, dass ich die vorliegende Arbeit in allen relevanten Teilen selbstständig durchgeführt habe.

Karoline Lang