



Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG



# MEZ Arbeitspapiere

Thorsten Klinger

## Die Messung berufsrelevanter Deutschkenntnisse in MEZ-2 mit Subskalen des WiDaF

MEZ-Arbeitspapier Nr. 12  
Hamburg, September 2021

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

Arbeitspapiere der Forschungsprojekte Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ) und Mehrsprachigkeit an der Schwelle zum Beruf (MEZ-2) an der Universität Hamburg

In der Reihe erscheinen Beiträge zu Themen, die den Arbeitsprozess des MEZ-Projekts betreffen. Die Beiträge stellen vorläufige Ergebnisse zu den theoretischen und empirischen Fragestellungen des Projekts sowie Lösungen zum technischen Umgang mit den Projektdaten vor.

Die in den *MEZ-Arbeitspapieren* vertretenen Meinungen sind die der Autor(inn)en und entsprechen nicht notwendigerweise den Auffassungen der Projektverantwortlichen.

**Principal Investigator MEZ-2-Projekt:**

Ingrid Gogolin, Universität Hamburg (Projektleitung)

**Bezug:**

[www.mez.uni-hamburg.de](http://www.mez.uni-hamburg.de)

**Kontakt:**

Mehrsprachigkeit an der Schwelle zum Beruf – MEZ-2

Universität Hamburg

Institut für Interkulturelle und International Vergleichende Erziehungswissenschaft

Postanschrift: Von-Melle-Park 8, 20146 Hamburg

Besucheranschrift: Alsterterrasse 1, 5. Stock, 20354 Hamburg

E-Mail: [mez@uni-hamburg.de](mailto:mez@uni-hamburg.de)

Tel.: +49 40 42838-2708

**Bitte zitieren Sie dieses Arbeitspapier wie folgt:**

Klinger, Thorsten (2021): Die Messung berufsrelevanter Deutschkenntnisse in MEZ-2 mit Subskalen des WiDaF. MEZ Arbeitspapier Nr. 12. Hamburg (Universität Hamburg). [www.mez.uni-hamburg.de](http://www.mez.uni-hamburg.de)

Thorsten Klinger

# Die Messung berufsrelevanter Deutschkenntnisse in MEZ-2 mit Subskalen des WiDaF

MEZ-Arbeitspapier Nr. 12

## Zusammenfassung:

Die Untersuchung MEZ-2 richtet sich auf Fragen nach der Sprachentwicklung an der Schwelle zum Beruf. Neben Testverfahren, die sich auf bildungsbezogene multiliterale Sprachfähigkeiten beziehen, waren für die in MEZ-2 in den Blick genommene Phase des Übergangs zusätzliche Tests vorgesehen, die beruflich relevante Sprachkompetenzen im Deutschen messen. Für receptive Fähigkeiten kamen in der ersten Welle der Studie Subtests des standardisierten Tests Deutsch als Fremdsprache in der Wirtschaft (WiDaF, Test 10) zum Einsatz, der von der Deutsch-Französischen Industrie- und Handelskammer entwickelt wurde (Riegler-Poyet et al. 2008). Der WiDaF richtet sich an Personen mit einer anderen Herkunftssprache als Deutsch, die auf fortgeschrittenem Niveau Deutsch als Fremdsprache lernen. Die Entscheidung für seine Erprobung im Rahmen der Untersuchung war dadurch begründet, dass es für berufsbezogene Sprachkenntnisse wenige fachlich übergreifende Kompetenzfeststellungsverfahren gibt.

In MEZ-2 wurden 27 Items aus dem Teil „Fachlexik“ („Synonyme“ und „Sprachbausteine“) verwendet. Die Erprobung der beiden Subtestmodule im Rahmen von MEZ-2 ergab, dass sie für die Stichprobe von Personen, die Deutsch als Erst- oder Zweitsprache in Deutschland erworben haben, sich als wenig geeignet erweist. Die Aufgaben waren tendenziell zu leicht für die Proband(inn)en. Die verwendeten Module „Fachlexik A und B“ messen bei unserer Stichprobe keine eindimensionalen Konstrukte, die einzelnen Aufgaben sind anhand unserer Daten nicht eindeutig benennbaren Kompetenzbereichen zuzuordnen. Die aus den Aufgaben errechneten Punktskoren erreichen keine befriedigende Reliabilität. Die Testmodule wurden für einen anderen Anwenderkreis als dem in MEZ-2 untersuchten entwickelt. Die Hoffnung, dass sich die gewählten Testmodule durch die Spezialisierung auf das „Wirtschaftsdeutsche“ für Kompetenzmessungen auch bei in Deutschland aufgewachsenen jungen Leuten am Übergang zum Beruf eignen würde, konnte nicht befriedigend bestätigt werden.

## Abstract:

The MEZ-2 study investigates language development in a transition to vocational life and the labour market. In addition to tests that focus on education-related multiliterate language skills, MEZ-2 included additional tests that measure occupationally relevant language proficiencies in German for the transition phase. For receptive skills, the first wave of the study used subtests of the standardized test "Deutsch als Fremdsprache in der Wirtschaft" (WiDaF, Test 10), which was developed by the Franco-German Chamber of Industry and Commerce (Riegler-Poyet et al. 2008). The WiDaF is aimed at individuals who are learning German as a foreign language at an advanced level. The decision to test the tool in the context of the MEZ-2-study was motivated by the fact that there are few cross-disciplinary competence assessment procedures for job-related language skills available. MEZ-2 included 27 items from the section "professional vocabulary".

The evaluation of the two subtest modules in MEZ-2 has shown that they are not applicable for the present sample of persons who have acquired German as a first or second language in Germany. The tasks tended to be too easy for the participants. Based on our data, the modules used do not measure unidimensional constructs, and the individual tasks cannot be clearly assigned to identifiable skill domains. The scores calculated from the tasks do not achieve satisfactory reliability. The test modules were developed for a different target group than the one investigated in MEZ-2. The expectation that, due to their specialization in "business German", the selected test modules measure the competences of young people living in Germany could not be confirmed satisfactorily.



## Inhaltsverzeichnis

1. Sprachentwicklung an der Schwelle zum Beruf: Die Erfassung berufsbildungs- und arbeitsmarktrelevanter sprachlicher Kenntnisse in MEZ-2.....	7
2. Die in MEZ-2 verwendeten Aufgaben des WiDaF .....	8
3. Zusammensetzung der Stichprobe.....	9
4. Prüfung der Skalenqualität.....	9
4.1 Itemschwierigkeit .....	9
4.2 Trennschärfe und interne Konsistenz der Items .....	10
4.3 Hauptkomponentenanalyse .....	11
5. Folgerungen für MEZ-2.....	13
Literaturverzeichnis.....	15



## 1. Sprachentwicklung an der Schwelle zum Beruf: Die Erfassung berufsbildungs- und arbeitsmarktrelevanter sprachlicher Kenntnisse in MEZ-2

Die Untersuchung MEZ-2 verfolgt die mehrsprachige Entwicklung von Jugendlichen mit und ohne Migrationshintergrund während der ersten Schritte in die berufliche Integration (Gogolin et al. 2021). Als Folgeuntersuchung der Studie „Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ)“ (Gogolin et al. 2017) basiert MEZ-2 auf Daten von Personen, die bereits an dieser vorangegangenen Studie teilgenommen haben. Es wird in der Phase des Übergangs in den Beruf eine Weiterentwicklung sprachlicher Fähigkeiten in Richtung auf berufliche bzw. akademische Fachsprachlichkeit angenommen, die sich vornehmlich auf die deutsche Sprache, eventuell auch auf die englische Sprache beziehen dürfte (Gogolin et al. 2021, 12). Neben den bislang schon in MEZ eingesetzten Instrumenten, die produktive, rezeptive und allgemeinsprachliche Kompetenzen messen, sollten in der Studie daher zusätzlich Tests Anwendung finden, die solche allgemeinen berufsbezogenen Sprachfähigkeiten messen, die in dieser Phase zunehmende Bedeutung erhalten. Die Tests sollten sich auf Qualifikationen beziehen, die für ein möglichst breites Spektrum von beruflichen Tätigkeiten relevant sind.

Eine entsprechende intensive Recherche ergab, dass in Bezug auf berufsrelevante deutschsprachige Lesefähigkeiten nur wenige geeignete Testverfahren zur Verfügung stehen (vgl. auch Ziegler et al. 2012; Redder et al. 2010). In den vorliegenden Studien zu Übergängen wurde zwar Lesefähigkeit als Indikator für Sprachkompetenzen verwendet (z.B. NEPS, IBIS, ULME II), jedoch wurde hierbei teils allgemeine Lesefähigkeit getestet, teils wurden Verfahren eingesetzt, die nicht in den zeitlichen Rahmen der Erhebung von MEZ-2 integrierbar waren. Der Lesetest LTB-3 (Drommler et al. 2006) war zunächst in Erwägung gezogen worden, der Einsatz wurde aber aufgrund eingeschränkter Validität (Jordan 2011) wieder verworfen. Die Suche nach einem anerkannten reliablen und validen Instrument, das weniger auf die Messung allgemeiner als speziell auf die Messung berufsrelevanter Lesefähigkeiten gerichtet ist, führte schließlich zu dem Test „Deutsch als Fremdsprache in der Wirtschaft – WiDaF“ (vgl. Deutsch-Französische Industrie- und Handelskammer o.J.)<sup>1</sup>. Dieser standardisierte berufsbezogene Deutschttest ist für Personen mit Deutsch als Fremdsprache konzipiert und misst die Fähigkeit, Wirtschaftsdeutsch am Arbeitsplatz einzusetzen (Bewertung des Hör- und Leseverständnisses unter spezieller Berücksichtigung der Grammatikkenntnisse). Der Test überprüft die Sprachniveaus von A2 bis C2 gemäß dem Gemeinsamen Europäischen Referenzrahmen. Er erfüllt nach Aussage der Herausgeber die testtheoretischen Kriterien der Validität, Reliabilität und Objektivität und unterliegt „regelmäßigen testmethodischen Überprüfungen und Analysen“ (<https://www.francoallemand.com/dienstleistungen/aus-und-weiterbildung/test-widaf-qualitaet>).<sup>2</sup> Für die Testung in MEZ-2 bot sich die Verwendung von Teilen des insgesamt 150 Minuten umfassenden Tests aus dem Abschnitt „Vokabular“ an. Inwieweit sich – wie im Fall von MEZ-2 – Testteile auch für den Einsatz bei deutschen

---

<sup>1</sup> Wir bedanken uns bei der Deutsch-Französischen Industrie- und Handelskammer für die Bereitstellung der Testteile und die Erlaubnis, sie im Rahmen der Studie verwenden zu dürfen.

<sup>2</sup> Die entsprechenden Belegzahlen konnten uns von den Herausgebern des Tests leider nicht zur Verfügung gestellt werden.

Bildungsinländern mit Deutsch als Erst- oder Zweitsprache eignen, ließ sich aufgrund des Zeitrahmens der Planungen im Vorfeld der MEZ-2-Erhebung mit kleinen Fallzahlen pilotieren, die Überprüfung ihrer Eignung konnte aber nicht hinreichend empirisch abgesichert werden. Wir entschieden uns dennoch für den probeweisen Einsatz im Rahmen der ersten Welle der MEZ-2-Online-Erhebung mit der Option, den Test auch in der zweiten Welle zu verwenden.

Der folgende Beitrag überprüft, inwieweit die Messung berufsbezogener Sprachfähigkeiten mit den verwendeten Modulen des WiDaF an der besonderen Stichprobe der MEZ-2-Proband(inn)en erfolgreich war.

## **2. Die in MEZ-2 verwendeten Aufgaben des WiDaF**

Der WiDaF ist ein standardisierter schriftlicher Test, der in autorisierten Testzentren abgelegt werden kann. Er dient in seiner Vollversion der Zertifizierung berufsbezogener Deutschkenntnisse gemäß den Niveaustufen des Gemeinsamen Europäischen Referenzrahmens (Riegler-Poyet et al. 2008). Der Gesamttest richtet sich auf Fachlexik, Grammatik, Lese- sowie auf Hörverständnis. Die Testaufgaben sind im Rahmen der originalen Testdurchführung papierbasiert im Multiple-Choice-Verfahren zu lösen.

Im Rahmen der Untersuchung MEZ-2 wurden verschiedene Testverfahren zur Bestimmung von Sprachkenntnissen in Deutsch, den Herkunftssprachen Russisch und Türkisch sowie in der Fremdsprache Englisch eingesetzt (vgl. Gogolin et al. 2021). Für die Testung berufsbezogener Sprachkenntnisse im Deutschen stand in diesem Kontext nur ein begrenzter zeitlicher Rahmen zur Verfügung, so dass die Durchführung des WiDaF-Gesamttests (150 Minuten) nicht in Frage kam. Aufgrund der Tatsache, dass die Studienteilnehmenden Schulen in Deutschland besucht haben und weitere Verfahren zu literalen Fähigkeiten im Lesen und Schreiben zum Einsatz kamen, waren wir insbesondere an Testteilen zum berufsrelevanten Fachwortschatz interessiert. Dementsprechend wurden 27 Items aus dem Teil „Fachlexik“ für die Untersuchung ausgewählt. Zur Anwendung kamen aus dem Testheft 10 jeweils die vollständigen Itemsätze der Aufgaben „Synonyme“ (10 Items) und „Sprachbausteine“ (17 Items).<sup>3</sup>

Für die Aufgabe „Synonyme“ werden zu einem vorgegebenen Einzelwort (6 Items) sowie zu einem unterstrichenen Wort innerhalb eines Satzes (4 Items) je vier alternative Wörter präsentiert, von denen jeweils eines die richtige Lösung, nämlich das Synonym, darstellt. Die Aufgabe „Sprachbausteine“ gibt pro Item einen Satz mit in beruflichen Kontexten üblichen Formulierungen wieder; jeder Satz enthält eine Lücke, für die aus ebenfalls vier Alternativen der passende Begriff auszuwählen ist. Für die Erhebung in MEZ-2 wurde der im Original papierbasierte Test in ein onlinebasiertes Format überführt. Für die Bearbeitung der 27 Items wurde ein Zeitlimit von 10 Minuten vorgegeben (IEA Hamburg 2020, 6ff.).

---

<sup>3</sup> Aus Gründen des Testschutzes können hier keine Beispielaufgaben und Lösungen wiedergegeben werden. Die im WiDaF-Trainingsbuch (Riegler-Poyet et al. 2008) publizierten Übungsaufgaben unterscheiden sich im Aufgabenformat von den in MEZ-2 verwendeten Items, so dass auch diese nicht zur Illustration der Aufgabenstellungen tauglich sind.

### 3. Zusammensetzung der Stichprobe

Am ersten Messzeitpunkt der MEZ-2-Studie wurden 728 Proband(inn)en erreicht, die mindestens an der ersten von zwei Testsitzungen teilnahmen (vgl. IEA Hamburg 2020; Gogolin et al. 2021). Die Anwendung des WiDaF erfolgte im Online-Modul der ersten Erhebungssitzung. Daran nahmen 416 Proband(inn)en teil, die monolingual mit der deutschen Sprache aufgewachsen sind (57,1%), sowie 140 Proband(inn)en mit russischer (19,2%) und 138 mit türkischer Herkunftssprache (19,0%); 34 Proband(inn)en (4,7%) konnten als mehrsprachig mit Deutsch und einer anderen Sprache klassifiziert werden.<sup>4</sup> Zwei Drittel der Teilnehmenden waren weiblich. Die Panelstudie MEZ wurde mit zwei Startkohorten durchgeführt, die in der ersten Welle Anfang 2016 die 7. bzw. 9. Schulklasse besuchten. In der ersten Welle der Folgeuntersuchung MEZ-2 gehörten noch 401 Teilnehmende (55,1%) der jüngeren Kohorte mit einem Durchschnittsalter von dann 17,4 Jahren an, 327 Personen (44,9%) zählten zur älteren Kohorte mit einem Durchschnittsalter von 19,6 Jahren. Der weitaus größte Teil (86 Prozent) der jüngeren Kohorte besuchte zum Testzeitpunkt eine allgemeinbildende Schule, während der entsprechende Anteil bei den Älteren lediglich 28 Prozent betrug. Bei Letzteren befand sich ein Viertel in einer dualen oder vollschulischen Ausbildung. Jeweils etwa 20 Prozent studierten oder machten derzeit „etwas anderes“ (überwiegend Freiwilligendienst, Jobben, Work & Travel).

### 4. Prüfung der Skalenqualität

Im ersten Schritt werden die Lösungshäufigkeiten der Items beschrieben. Anschließend wird als Reliabilitätsmaß die interne Konsistenz der Items geprüft. Dabei werden die auf die jeweilige Aufgabe „Synonyme“ (10 Items) und „Sprachbausteine“ (17 Items) bezogenen Items sowohl für jedes Itemset getrennt als auch gemeinsam als Gesamtskala analysiert. Mittels einer Hauptkomponentenanalyse wird anschließend explorativ überprüft, inwieweit die verwendeten Items eine oder mehrere Dimensionen berufsrelevanter Wortschatzphänomene abbilden.

#### 4.1 Itemschwierigkeit

Die Itemschwierigkeit wird in Tabelle 1 (Spalte 2) mit den Prozentangaben der korrekten Lösungen angegeben. Von den insgesamt 27 Items konnten 17 von mehr als 90 Prozent der Teilnehmenden richtig beantwortet werden. Damit waren annähernd zwei Drittel der Multiple-Choice-Aufgaben für die Teilnehmenden deutlich zu leicht. Große Schwierigkeiten bereitete die Beantwortung des Items 1 und insbesondere des Items 23, die nur von etwa 28 Prozent bzw. 16 Prozent der Befragten korrekt gelöst wurden. Für die in MEZ-2 getestete Befragtengruppe muss der Test damit als zu leicht mit einer deutlichen Tendenz zu Deckeneffekten charakterisiert werden.

---

<sup>4</sup> Die revidierte Klassifizierung wurde nachträglich anhand von Fragebogenangaben vorgenommen. Für die Organisation der Feldarbeit der IEA war eine davon abweichende Einteilung maßgeblich (Brandt et al. 2019). Aus diesem Grund weichen die hier berichteten Zahlen von denen im Methodenbericht der IEA ab (vgl. IEA Hamburg 2020, 12f.).

#### 4.2 Trennschärfe und interne Konsistenz der Items

Die Trennschärfe eines Items gibt an, inwieweit die Lösung des Ergebnisses einer Skala mit der Beantwortung dieses einzelnen Items vorhersagbar ist (Bortz und Döring 2006). Tabelle 1 gibt in den Spalten 3 bis 5 die Trennschärfe-Kennwerte, die interne Konsistenz für die Subskalen sowie für die aus beiden gebildete Gesamtskala wieder. Bei den zehn Items der Skala „Synonyme“ erreicht lediglich eine Item-Skala-Korrelation den kritischen Wert von 0,3. Die übrigen Items weisen in Bezug auf die Subskala durchweg keine hinreichende Trennschärfe auf. Die interne Konsistenz dieser Subskala ist mit einem Cronbachs  $\alpha$  von 0,44 als inakzeptabel zu bewerten. Bei der Subskala „Sprachbausteine“ liegen vier von 17 Items knapp über einer Item-Skala-Korrelation von 0,3, während 13 Items schwächer mit der Skala korrelieren. Mit einem  $\alpha$  von 0,52 ist auch hier die interne Konsistenz zu niedrig.

Tabelle 1: Trennschärfe und interne Konsistenz der verwendeten WiDaF-Items nach Subskalen (n=728)

<i>Item</i>	% korrekt gelöst	Trennschärfe* Subskala 1 (Synonyme)	Trennschärfe* Subskala 2 (Sprachbausteine)	Trennschärfe* Gesamtskala (Subskalen 1 + 2)
(1)	(2)	(3)	(4)	(5)
1	27,9	0,190		0,219
2	99,2	0,143		0,090
3	97,5	0,178		0,163
4	68,8	0,219		0,265
5	72,5	<b>0,301</b>		<b>0,375</b>
6	93,7	0,279		<b>0,307</b>
7	98,5	0,150		0,191
8	96,3	0,055		0,107
9	98,2	0,111		0,183
10	94,4	0,265		0,298
11	58,2		0,176	0,220
12	96,6		0,282	<b>0,362</b>
13	78,8		0,202	0,210
14	92,8		0,189	0,200
15	98,8		<b>0,304</b>	<b>0,321</b>
16	95,0		0,232	0,244
17	95,6		0,241	0,283
18	98,8		<b>0,304</b>	0,286
19	93,1		0,234	0,246
20	95,9		<b>0,332</b>	<b>0,323</b>
21	39,1		0,069	0,105
22	98,3		0,212	0,180
23	16,4		0,027	0,034
24	63,5		0,100	0,127
25	81,3		0,209	0,257
26	88,3		<b>0,371</b>	<b>0,414</b>
27	96,7		0,269	0,251
<b>Cronbachs <math>\alpha</math></b>		<b>0,44</b>	<b>0,52</b>	<b>0,65</b>

\* korrigierte Item-Skala-Korrelation

Für die aus allen 27 Items gebildete Skala ergibt sich mit einem Cronbachs  $\alpha$  von 0,65 ein geringfügig besseres Ergebnis, das aber auf die höhere Itemzahl zurückgeht, von der dieser Kennwert deutlich abhängig ist (Bortz und Döring 2006). Die Trennschärfe der Items ist auch bei der Gesamtskala generell zu gering. Eine nach Sprachgruppen getrennte Berechnung der internen Konsistenz kommt zu vergleichbar niedrigen Werten.<sup>5</sup>

### 4.3 Hauptkomponentenanalyse

Cronbachs  $\alpha$  zielt auf geringe itemspezifische Varianz in einem eindimensionalen Test ab, stellt aber keinen Nachweis der Eindimensionalität eines Tests oder einer Skala dar (Taber 2018; Cortina 1993, 103). Die hier vorgefundenen geringen Trennschärfen und niedrigen  $\alpha$ -Koeffizienten geben bereits Hinweise darauf, dass die verwendeten Items in unserer Zielgruppe kein homogenes Konstrukt abbilden. Zur Überprüfung wird hier ergänzend eine Hauptkomponentenanalyse (PCA) durchgeführt, die die Dimensionalität von Daten durch ihre Zerlegung in Hauptkomponenten erkundet (Jolliffe und Cadima 2016).

Die Hauptkomponentenanalyse wird über alle 27 Items durchgeführt. Das Kaiser-Meyer-Olkin-Kriterium wird zur Bestimmung der Stichprobeneignung für Hauptkomponenten- und Faktorenanalysen herangezogen, wobei zumeist Werte ab 0,6 als annehmbar gelten (Mayer 2018, 8). Für unsere Stichprobe wird ein akzeptabler Wert von 0,786 erreicht. Die 27 Iteminformationen lassen sich nach dem Kaiser-Kriterium (Eigenwert  $\geq 1$ ) auf eine Lösung mit acht Komponenten reduzieren, die zusammen 46,91 Prozent der Gesamtvarianz erklären (Tabelle 2). Eine Sichtung des Scree-Plots ließe eine Reduzierung auf 3 Komponenten zu, die aber auf eine Varianzerklärung von 26 Prozent käme und daher verworfen wurde, zumal diese Lösung keine Verbesserung der Interpretation der Komponenten erbrachte. Eine Varimax-Rotation mit Kaiser-Normalisierung konvergierte in 35 Iterationen. Die sich daraus ergebenden Komponentenladungen sind in Tabelle 3 aufgeführt.

Tabelle 2: Hauptkomponentenanalyse, Eigenwerte und erklärte Varianz (rotierte Summe der quadrierten Ladungen)

Komponente	Eigenwerte	erklärte Varianz in %
1	2,164	8,014
2	2,124	7,865
3	1,813	6,714
4	1,593	5,900
5	1,360	5,038
6	1,335	4,944
7	1,162	4,303
8	1,115	4,128

<sup>5</sup> Monolingual-Dtsch.:  $\alpha=0,62$ ; Dtsch.-Russ.:  $\alpha=0,53$ ; Dtsch.-Trk.:  $\alpha=0,67$ ; anders mehrsprg.:  $\alpha=0,42$ .

Messung berufsrelevanter Deutschkenntnisse in MEZ-2

Tabelle 3: Hauptkomponentenanalyse, rotierte Komponentenladungen der Items (Varimax)

Item	Rotierte Komponentenladungen							
	1	2	3	4	5	6	7	8
1	0,094	-0,153	0,103	0,013	<b>0,594</b>	0,255	0,019	0,098
2	-0,104	-0,018	0,147	<b>0,715</b>	0,011	0,005	-0,046	-0,145
3	0,118	-0,101	0,056	<b>0,701</b>	0,025	-0,008	0,077	-0,008
4	0,150	0,110	-0,091	0,185	0,171	<b>0,458</b>	0,091	0,015
5	<b>0,611</b>	-0,067	0,052	0,055	0,187	0,173	0,015	0,080
6	0,253	0,168	0,040	<b>0,492</b>	0,067	0,117	0,110	0,059
7	0,086	0,017	<b>0,643</b>	0,195	0,017	-0,067	0,020	0,138
8	0,087	0,163	-0,113	-0,010	0,186	-0,047	<b>-0,525</b>	0,133
9	0,061	0,115	<b>0,723</b>	0,002	-0,005	-0,007	0,132	-0,102
10	0,596	0,005	0,328	-0,051	-0,051	0,022	-0,123	-0,033
11	0,118	0,024	0,060	-0,045	0,043	<b>0,709</b>	0,004	0,043
12	<b>0,495</b>	0,087	0,229	0,308	-0,029	-0,013	-0,067	0,262
13	0,096	0,114	-0,083	0,018	<b>0,649</b>	-0,007	-0,012	-0,044
14	-0,113	0,378	0,195	0,088	<b>0,405</b>	-0,011	-0,329	-0,012
15	0,133	<b>0,447</b>	<b>0,565</b>	0,064	0,012	-0,035	-0,067	-0,048
16	0,054	0,303	<b>0,402</b>	0,055	0,025	0,136	-0,093	0,073
17	0,479	0,244	-0,109	0,297	-0,207	0,008	-0,154	0,169
18	0,105	<b>0,615</b>	0,216	-0,074	0,047	0,023	0,104	-0,045
19	0,296	0,340	-0,190	-0,052	0,007	0,305	0,107	-0,141
20	0,270	<b>0,596</b>	0,032	0,074	-0,054	0,019	-0,087	0,004
21	0,327	0,066	-0,010	-0,041	0,306	<b>-0,472</b>	0,368	-0,145
22	-0,131	<b>0,628</b>	0,162	-0,049	0,004	0,100	0,044	0,043
23	-0,135	0,133	-0,040	0,143	0,111	0,031	<b>0,663</b>	0,197
24	0,134	0,031	0,043	-0,109	0,071	0,104	0,064	<b>0,822</b>
25	0,362	0,052	0,027	-0,017	0,147	0,359	0,126	-0,374
26	<b>0,547</b>	0,207	0,011	0,069	0,198	0,103	-0,066	-0,051
27	0,063	<b>0,421</b>	-0,002	0,118	0,303	-0,095	-0,208	0,125

Faktorladungen  $\geq 0,4$  sind fett hervorgehoben

Eine inhaltliche Interpretation der Komponenten auf Basis der Ladungen gestaltet sich als schwierig, zumal auf je einer Komponente nur wenige Items laden, die Ladungen unsystematisch über die beiden Subskalen streuen und nur vereinzelt hohe Faktorladungen zu beobachten sind.

Die Hauptkomponentenanalyse bestätigt somit den Eindruck, dass in der MEZ-2-Stichprobe, die aus Bildungsinländern zusammengesetzt ist, mit den gewählten Items kein homogenes Konstrukt abgebildet wird.

## **5. Folgerungen für MEZ-2**

Der Test WiDaF richtet sich ausdrücklich an Personen mit einer anderen Herkunftssprache als Deutsch, die auf fortgeschrittenem Niveau Deutsch als Fremdsprache lernen. Für diesen Personenkreis wurde er entwickelt und auf diesen beziehen sich die Aussagen der Entwickler in ihrer Bewertung der Objektivität, Validität und Reliabilität seiner Aufgaben und der mit ihnen erzielten Ergebnisse.

Die Untersuchung MEZ-2 bezieht sich auf eine andere Grundgesamtheit: Es geht um junge Menschen, die in Deutschland ihre Schulbildung erworben und teilweise bereits abgeschlossen haben und deren Familiensprache Deutsch und zum Teil zusätzlich Russisch, Türkisch sowie ggfs. eine weitere Sprache ist. Dass der Test unsererseits in Erwägung gezogen wurde, um die wirtschafts- und berufsbezogenen Sprachkenntnisse unserer Proband(inn)en einzuschätzen, ist einerseits auf einen Mangel an hierauf spezialisierten Verfahren zurückzuführen. Andererseits erschien es uns als aussichtsreich, dass der Test durch seinen Bezug auf die Besonderheiten der in Wirtschaft und Beruf geltenden Lexik und Formulierungen auch eine hinreichende Varianz entsprechender Kenntnisse bei beruflich wenig erfahrenen jungen Leuten abbildet, deren Schulbildung andauert oder noch nicht lange zurückliegt.

Die hier präsentierten Ergebnisse zeigen, dass der von uns gewählte Ausschnitt des WiDaF-Tests für unsere Zielgruppe wenig geeignet ist. Die Aufgaben sind überwiegend zu leicht und erreichen keine ausreichende Trennschärfe, so dass ihre Beantwortung keinen systematischen Aufschluss darüber gibt, wie gut die Teilnehmenden an unserer Untersuchung berufsbezogene Lexik oder Teildomänen davon beherrschen. Summenskalen auf Basis der erzielten Punkte sind schwer interpretierbar, da aufgrund der geringen Konsistenz und der fehlenden Homogenität der zugrundeliegenden Testitems unklar bleibt, was damit eigentlich gemessen wird. Aus diesen Gründen wurde auf einen weiteren Einsatz der WiDaF-Komponenten im Rahmen unserer Untersuchung verzichtet.

Unsere Ergebnisse stellen die Eignung der untersuchten Testteile und des Gesamttests WiDaF für die Personen nicht in Frage, für die der Test konstruiert wurde und für die er üblicherweise verwendet wird. Auch unsere Heraustrennung von Testteilen aus dem Gesamtzusammenhang ist hier zu erwähnen, da die Bewertung der Sprachkompetenzen einer Person in der regelhaften Testpraxis des WiDaF auf den Ergebnissen aller seiner Teile erfolgt.

Als Alternative im Rahmen unserer Studie bot sich zwischenzeitlich die Gelegenheit für die Nutzung einer eigens für MEZ-2 neu erstellten online-basierten Version des Tests „MaK-Adapt Lesen“ (vgl. Ziegler et al. 2012). Dieser Test für berufsbezogenes Leseverständnis im Deutschen basiert auf einem theoretischen Modell, das speziell auf die Messung von Leseanforderungen im beruflichen Handlungskontext ausgerichtet ist (FDZ Bildung o.J.). Die Version konnte rechtzeitig für die zweite Erhebungswelle von MEZ-2 zur Verfügung gestellt werden, Ergebnisse werden in einem späteren Band dieser Reihe vorgestellt.

## Literaturverzeichnis

Bortz, Jürgen; Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.

Brandt, Hanne; Dünkel, Nora; Heimler, Julia (2019): *Konstrukt und Operationalisierung lebensweltlicher Ein- und Mehrsprachigkeit im Rahmen des Projekts Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ)*. MEZ Arbeitspapier Nr. 4, Hamburg (Universität Hamburg), [www.mez.uni-hamburg.de](http://www.mez.uni-hamburg.de)

Cortina, Jose M. (1993): What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. doi:10.1037/0021-9010.78.1.98.

Deutsch-Französischen Industrie- und Handelskammer (o.J.): <https://www.francoallemand.com/dienstleistungen/aus-und-weiterbildung/test-widaf/>

Drommler, Rebecca; Linnemann, Markus; Becker-Mrotzek, Michael; Haider, Hilde; Stevens, Tobias; Wahlers, Judith (2006): *Lesetest für Berufsschüler/innen LTB-3*. Handbuch. Duisburg: Gilles und Francke Verlag.

FDZ Bildung (Forschungsdatenzentrum Bildung) (o.J.): *Testinstrument: MaK-adapt-Test Lesen*. DOI: 10.7477/259:184:40

Gogolin, Ingrid; Klinger Thorsten; Lagemann, Marina; Schnoor, Birger in Zusammenarbeit mit Christoph Gabriel, Michel Knigge, Marion Krause und Peter Siemund (2017): *Indikation, Konzeption und Untersuchungsdesign des Projekts Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ)*. MEZ Arbeitspapier Nr. 1. Hamburg (Universität Hamburg), [www.mez.uni-hamburg.de](http://www.mez.uni-hamburg.de)

Gogolin, Ingrid; Klinger, Thorsten; Schnoor, Birger; Usanova, Irina (2021): *Mehrsprachigkeit an der Schwelle zum Beruf. Die Funktion sprachlicher Fähigkeiten für Berufsqualifizierung und Berufseingangs-mündung von Jugendlichen mit und ohne Migrationshintergrund (MEZ-2)*. MEZ Arbeitspapier Nr. 10. Hamburg (Universität Hamburg): [www.mez.uni-hamburg.de](http://www.mez.uni-hamburg.de)

IEA Hamburg (2020): *Feld- und Methodenbericht der IEA Hamburg für das Projekt „Mehrsprachigkeit an der Schwelle zum Beruf“ (MEZ-2) 1. Messzeitpunkt*. Bericht für die MEZ-2-Studienleitung. Hamburg: IEA Hamburg. <https://www.mez.uni-hamburg.de/bilder/pdf/mez-2-methodenbericht1-iea-pdf.pdf>

Jolliffe, Ian T., & Cadima, Jorge (2016): Principal component analysis: a review and recent developments. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>

Jordan, Roland (2011): *Entwicklung und Validierung eines Testverfahrens zur Ermittlung der Lesekompetenz und des mathematischen Textverständnisses mit empirischer Untersuchung an allgemeinbildenden und berufsbildenden Schulen*. Münster: Verlag für Wissenschaftliche Texte und Medien.

Mayer Sabrina J. (2018). Hauptkomponentenanalyse und explorative Faktorenanalyse. In: Wagemann C., Goerres A., Siewert M. (Hg.) Handbuch Methoden der Politikwissenschaft. Springer Reference Sozialwissenschaften. Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-16937-4\\_31-1](https://doi.org/10.1007/978-3-658-16937-4_31-1)

Redder, Angelika; Schwippert, Knut; Hasselhorn, Marcus; Forschner, S.; Fickermann, Detlef; Ehlich, Konrad (2010): Grundzüge eines nationalen Forschungsprogramms zu Sprachdiagnostik und Sprachförderung. ZUSE-Diskussionspapier Nr. 1.

Riegler-Poyet, Margarete; Straub, Bernard; Thiele, Paul (2008): *Das Testbuch Wirtschaftsdeutsch - Training zum Test WiDaF* (Neubearbeitung). Berlin und München: Langenscheidt.

Taber, Keith S. (2018): The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Res Sci Educ* 48, 1273–1296 (2018). <https://doi.org/10.1007/s11165-016-9602-2>

Ziegler, Birgit; Balkenhol, Aileen; Keimes, Christina; Rexing, Volker (2012): Funktionen und Erträge pädagogischer Diagnostik im wirtschafts- und berufspädagogischen Bereich. In: *bwp@* Ausgabe Nr. 22, Juni 2012. [http://www.bwpat.de/ausgabe22/ziegler\\_etal\\_bwpat22.pdf](http://www.bwpat.de/ausgabe22/ziegler_etal_bwpat22.pdf)