

Elektronisches Supplement 5. Beschreibung des Erhebungsablauf und Testdesign.

Die Schülerinnen und Schülern (SuS) wurden im Klassenverband in ihrem Klassenraum in zwei aufeinanderfolgenden Schulstunden (90 Minuten) an einem Vormittag getestet. In den ersten 45 Minuten bearbeiteten die SuS die ihnen zufällig zugewiesenen Testhefte von „FALKE“ an den bereitgestellten Notebooks. Vor Beginn der Testung wurde durch eine geschulte Testleitung das Programm für die Erhebung vorgestellt. Die Instruktion für die SuS umfasste Informationen zu „Flip dem Falken“, einem Vogel der Rätsel für SuS mitgebracht hat. Die aufgabenbezogene Instruktion erfolgte durch Beispiele auf den Notebooks. In den Beispielen zeigte „Flip“ den SuS, wie ein jeweiliges Rätsel zu lösen ist. Die SuS konnten die Beispiele in ihrer eigenen Geschwindigkeit anschauen und auch während der Testung immer wieder darauf zurückgreifen. Während des gesamten Testzeitraums waren zusätzlich zwei Testleitungen anwesend.

Aufgrund der erheblichen Unterschiede in der Lesegeschwindigkeit der SuS wurde den SuS, die zügig die Items bearbeiteten, danach die Möglichkeit gegeben, sich mit einem Ausmalbild von Flip zu beschäftigen, damit die übrigen SuS ruhig weiterarbeiten konnten. Die maximale Bearbeitungszeit für alle SuS betrug 30 Minuten. In der zweiten Hälfte der Testung wurden die Fragebögen und die schriftlichen Tests bearbeitet.

Die Testhefte der dritten Klassenstufen (K3) bestanden jeweils aus zwei Itemblöcken á circa neun Items und damit insgesamt circa 18 Items. Die Testhefte der vierten Klassenstufe (K4) bestanden jeweils aus drei Itemblöcken (siehe Tabelle 1) á circa 10 Items und damit aus circa 30 Items. Die unterschiedliche Länge der Testhefte in K3 und K4 wurde aufgrund der höheren Lesegeschwindigkeit der SuS in K4 angelegt, um die Itembearbeitungszeit von 30 Minuten jeweils auszunutzen. Die Itemblöcke wurden so rotiert, dass sie einmal an jeder Position der Testhefte platziert waren, damit sich mögliche Reihenfolgeeffekte bei der Schätzung der Itemparameter ausgleichen.

In der ersten Erhebung (E1) bestand der Itempool aus den Items, für die Testhefte gewählt wurden, aus insgesamt 96 Items (siehe Tabelle 2). Ursprünglich bestanden der Itempool aus 99 Items und die Itemblöcke aus jeweils aus neun Items. Drei Items in drei verschiedenen Blöcken (I4, I5 und I9) konnten aufgrund von technischen Darstellungs- (2x) und Speicherfehlern (1x) nicht mehr ausgewertet werden. Für die zweite Erhebung wurden die Items aus E1 erneut verwendet. Anhand der empirischen Schwierigkeiten wurden die Items in E1 auf die K3 und K4 der zweiten Erhebung (E2) verteilt. Um das Linking zwischen den Studien zu ermöglichen, wurden die Itemblöcke I1, I2, und I3 in identischer Form in allen drei Teilstudien verwendet. Eine Ausnahme hiervon ist I2. Der Itemblock I2 enthielt in der E1 K3 ein Item weniger als in E1 K3 und E2 K4, weil das letzte Item des Blocks sich in E1 als zu schwierig für die K3 herausstellte und daher nur noch in E2 K4 eingesetzt wurde. Für die E2 wurden der Itempool um 79 Items erweitert. Diese Items wurden nach Einschätzung der Itemautorinnen und -autoren auf die Testhefte der K3 und K4 verteilt. Um die Items in den drei Teilstichproben gemeinsam skalieren zu können, wurde auf einen hinreichenden Anteil an Linkitems geachtet. Es ergaben sich Anteile von E1 K3 zu E2 K3: 43 %, E1 K3 zu E2 K4: 47 % und E2 K3 zu E2 K3: 51 %;

Tabelle 1. Testdesign der Erhebungen 1 (E1) und 2 (E2) und der Klassenstufe 3 (K3) und 4 (K4). Die Itemblöcke (I) werden an unterschiedlichen Position (Pos) und in unterschiedlichen Testheften (TH) eingesetzt.

E1 K3				E2 K3				E2 K4				
TH	Pos		Items	TH	Pos		Items	TH	Pos			Items
	1	2			1	2			1	2	3	
1	I1 ⁽⁹⁾	I2 ⁽⁹⁾	18	1	I3 ^{*(9)}	I2 ^{*(8)}	17	1	I01 ^{*(9)}	I04 ⁽¹⁰⁾	I05 ⁽¹⁰⁾	29
2	I2 ⁽⁹⁾	I3 ⁽⁹⁾	18	2	I2 ^{*(8)}	I4 ⁽¹⁰⁾	18	2	I10 ⁽¹⁰⁾	I01 ^{*(9)}	I02 ^{*(9)}	28
3	I3 ⁽⁹⁾	I4 ⁽⁸⁾	17	3	I4 ⁽¹⁰⁾	I3 ^{*(9)}	19	3	I02 ^{*(9)}	I03 ^{*(9)}	I07 ⁽¹⁰⁾	28
4	I4 ⁽⁸⁾	I5 ⁽⁸⁾	16	4	I1 ^{*(9)}	I5 ⁽¹⁰⁾	19	4	I11 ⁽¹⁰⁾	I02 ^{*(9)}	I03 ^{*(9)}	28
5	I5 ⁽⁸⁾	I6 ⁽⁹⁾	17	5	I5 ⁽¹⁰⁾	I1 ^{*(9)}	19	5	I03 ^{*(9)}	I08 ⁽¹⁰⁾	I09 ⁽¹⁰⁾	29
6	I6 ⁽⁹⁾	I7 ⁽⁹⁾	18	6	I6 ⁽¹⁰⁾	I7 ⁽¹⁰⁾	20	6	I12 ⁽¹⁰⁾	I06 ⁽¹⁰⁾	I04 ⁽¹⁰⁾	30
7	I7 ⁽⁹⁾	I8 ⁽⁹⁾	18	7	I7 ⁽¹⁰⁾	I8 ⁽¹⁰⁾	20	7	I13 ⁽¹⁰⁾	I09 ⁽¹⁰⁾	I15 ⁽¹⁰⁾	30
8	I8 ⁽⁹⁾	I9 ⁽⁹⁾	18	8	I8 ⁽¹⁰⁾	I6 ⁽¹⁰⁾	20	8	I05 ⁽¹⁰⁾	I11 ⁽¹⁰⁾	I06 ⁽¹⁰⁾	30
9	I9 ⁽⁹⁾	I10 ⁽⁸⁾	17	Itemblöcke enthalten 10 Items, die *Itemblöcke sind aus E1 abzüglich Items mit ungünstigen Messeigenschaften.				9	I07 ⁽¹⁰⁾	I12 ⁽¹⁰⁾	I08 ⁽¹⁰⁾	30
10	I10 ⁽⁸⁾	I11 ⁽⁹⁾	17					10	I09 ⁽¹⁰⁾	I13 ⁽¹⁰⁾	I14 ⁽¹⁰⁾	30
Itemblöcke (I) enthalten jeweils 9 Items								11	I15 ⁽¹⁰⁾	I10 ⁽¹⁰⁾	I12 ⁽¹⁰⁾	30
* Drei Items wurden aufgrund technischer Probleme nicht ausgespielt. Vorgesehen waren 18.								12	I08 ⁽¹⁰⁾	I07 ⁽¹⁰⁾	I04 ⁽¹⁰⁾	30
								13	I06 ⁽¹⁰⁾	I14 ⁽¹⁰⁾	I13 ⁽¹⁰⁾	30
								14	I14 ⁽¹⁰⁾	I05 ⁽¹⁰⁾	I10 ⁽¹⁰⁾	30
								15	I01 ^{*(9)}	I15 ⁽¹⁰⁾	I11 ⁽¹⁰⁾	29

Tabelle 2. Überlappende und vereinigte Itemsets der Erhebungen 1 (E1) und 2 (E2) und der Klassenstufe 3 (K3) und 4 (K4).

	E1 K3	E2 K3	E2 K4	Union
E1 K3	96	139	174	
E2 K3	33	76	148	175*
E2 K4	69	75	147	

Anmerkungen: Anzahl Items überlappend zwischen Studien in Dunkelgrau. Anzahl Items der vereinigen Menge in hellgrau. *Anzahl Items der vereinigen Menge über alle drei Studien.