

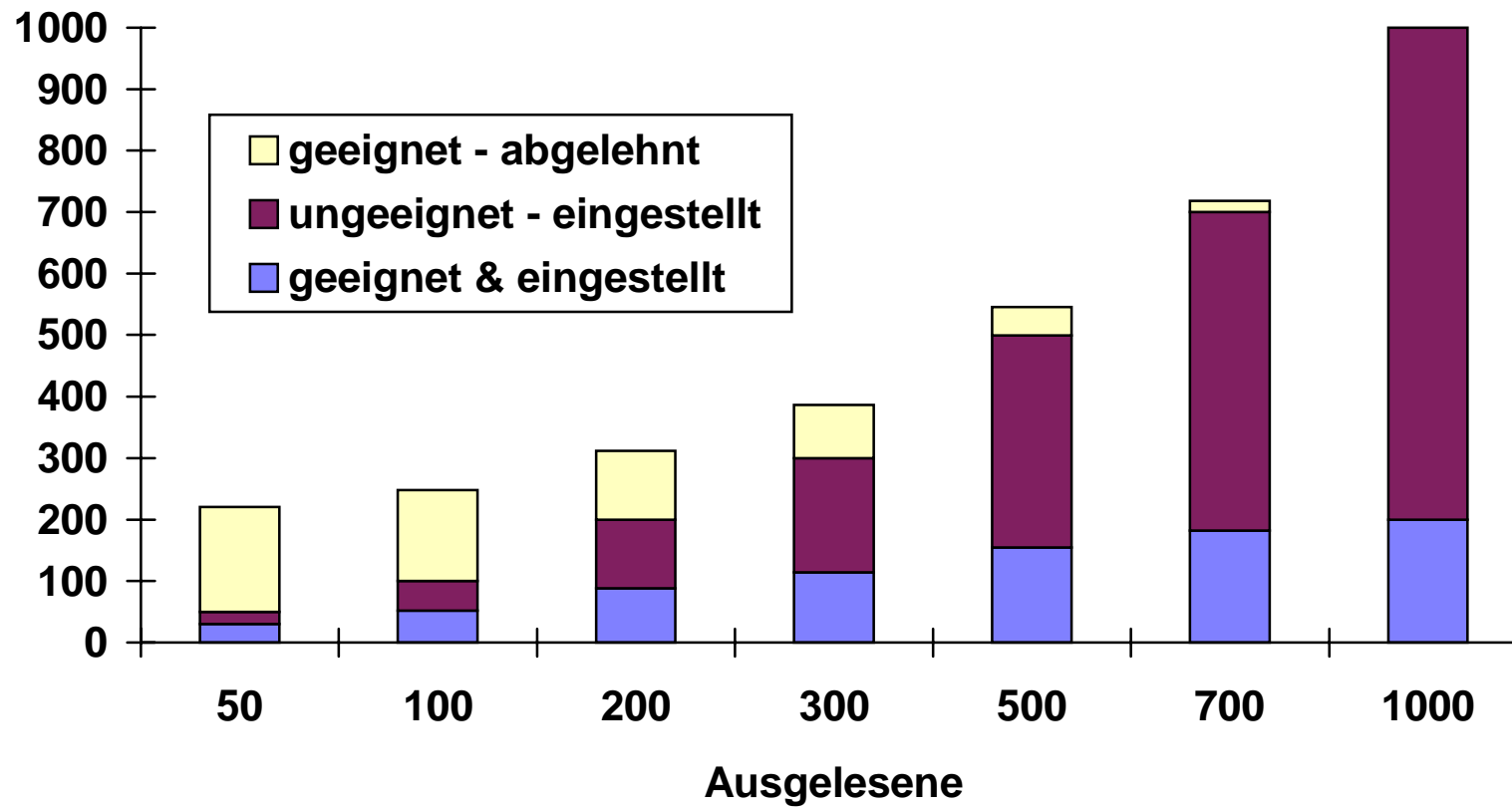
VERBESSERUNG VON ENTSCHEIDUNGEN DURCH TESTS

Diagnostische Va- lenz	0,00		0,30			0,50			0,70			
	+		-		+		-		+		-	
Eignung	+		-		+		-		+		-	
Ausgelesene	E	NE	E	E	NE	E	E	NE	E	E	NE	E
50	10	190	40	20	180	30	30	170	20	40	160	10
100	20	180	80	37	163	63	52	148	48	69	131	31
200	40	160	160	66	134	134	88	112	112	112	88	88
300	60	140	240	90	110	210	114	86	186	144	56	156
500	100	100	400	135	65	365	155	45	345	180	20	320
700	140	60	560	168	32	532	182	18	518	196	4	504
1000	200	0	800	200	0	800	200	0	800	200	0	800

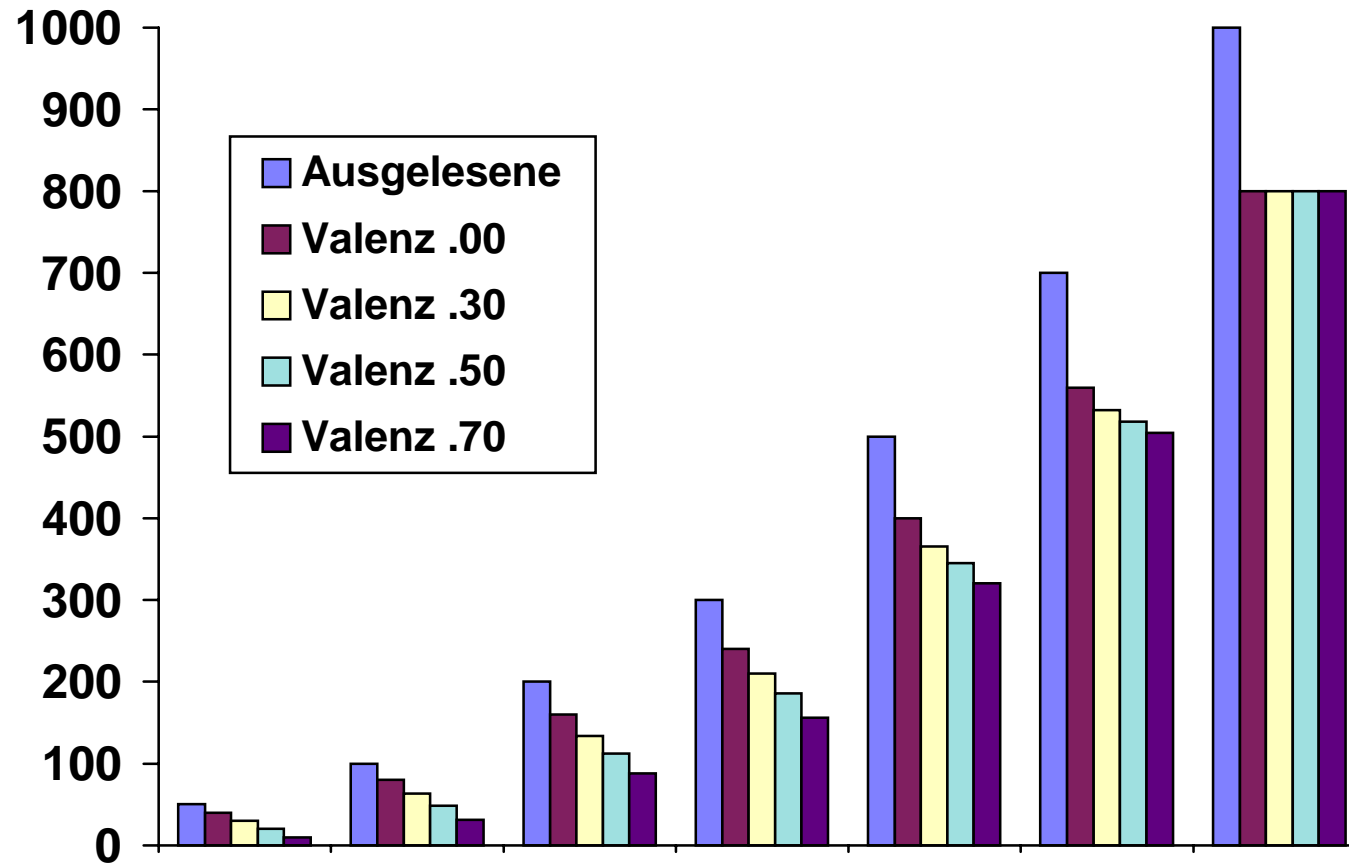
Tabelle nach H. C. Taylor & J. T. Russell (1939)

In den mit den Buchstaben E und NE überschriebenen Spalten finden sich die Anzahlen der Probanden, die Eingestellt bzw. Nicht eingestellt werden. Zugrunde gelegt wurde die Annahme, daß sich in einer unausgelesenen Population von 1000 Berufsanwärtern 200 geeignete Personen (= 20 %) befinden. Wählen wir nunmehr an Hand eines Tests, der z.B. eine diagnostische Valenz [= Gültigkeit] von $r_{xy} = 0,50$ besitzt [das ist eine nur mit einigem Aufwand erreichbare mittlere Übereinstimmung bzw. Vorhersage], 200 Personen aus, und zwar natürlich diejenigen von den Probanden, die in dem Test am besten abschneiden, so ergibt sich das folgende Bild: Unsere ausgelesene Gruppe enthält 88 Geeignete und 112 Ungeeignete (= 44 % Geeignete); andererseits befinden sich aber unter den zurückgewiesenen 800 Probanden ebenfalls noch 112 Geeignete (= 14 % der Zurückgewiesenen). In der Tat haben wir mehr Geeignete abgelehnt als angenommen. Besitzt unser Test eine diagnostische Valenz von $r_{xy} = 0,70$, so erlaubt er uns die Bildung einer ausgelesenen Gruppe von 200 Personen, in der sich der Prozentsatz der Geeigneten auf 56 % beläuft, wonach immer noch unter den Zurückgewiesenen 11 Prozent Geeignete verbleiben. Dieses Ergebnis bleibt weit hinter der Idealvorstellung von der Leistungsfähigkeit eines 'guten' Tests zurück; es ist jedoch mit dem einer aufs Geratewohl ($r_{xy} = 0,00$) erfolgten Zulassung zu vergleichen, bei der nur 40 Geeignete, aber 160 Ungeeignete in die ausgelesene Gruppe kämen, und daher 160 Geeignete abgelehnt würden. Der Prozentsatz der Geeigneten beliefe sich in der ausgewählten Gruppe auf bloß 20 Prozent. Sämtliche 200 Geeignete sind nur dann zu erfassen, wenn man überhaupt auf eine Auswahl verzichtet und alle 1000 Bewerber, damit freilich auch 800 Ungeeignete, einstellt. (Hofstädter, Einf. Psych., p.293)

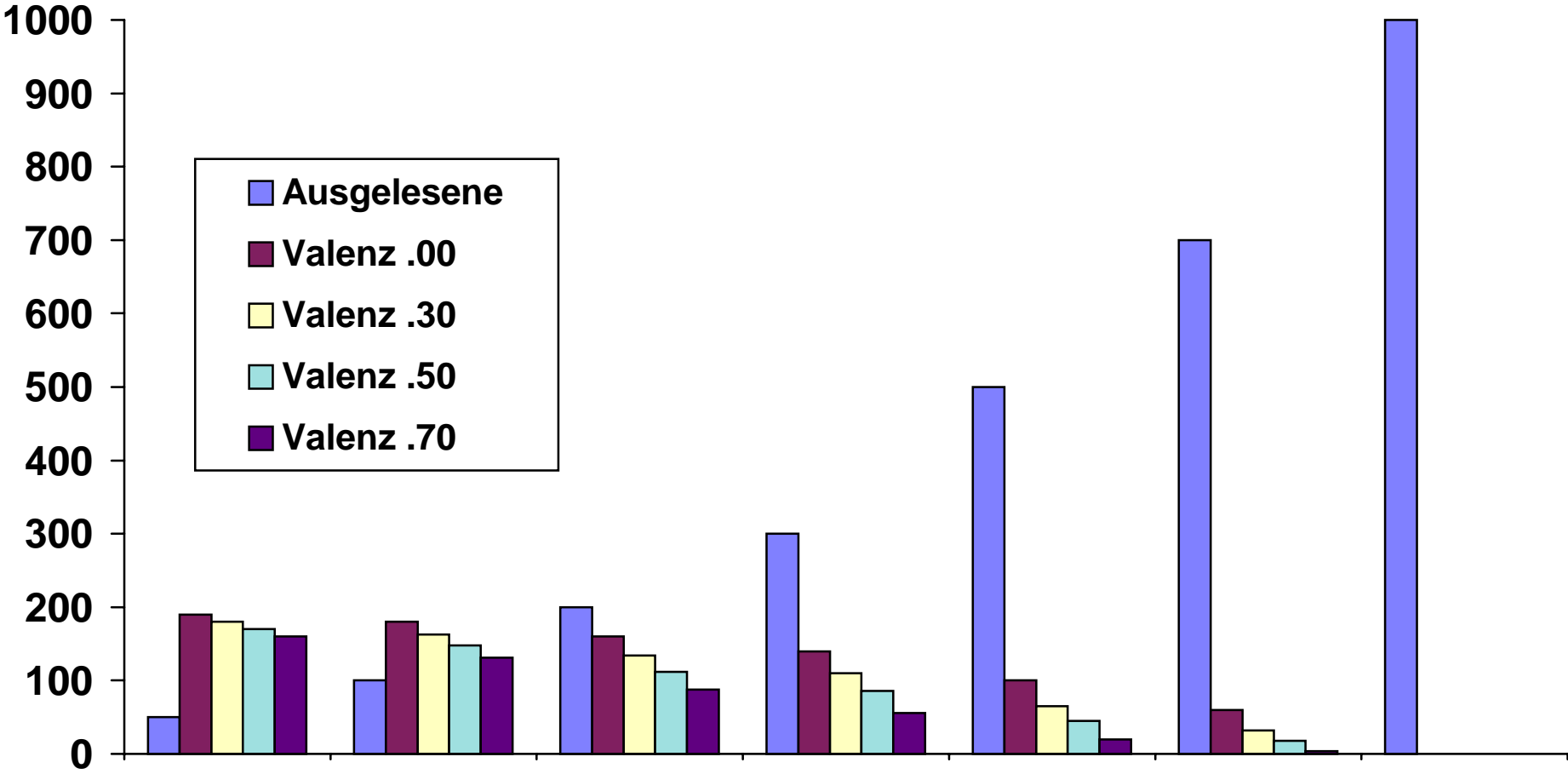
Einstellung und Eignung bei Diagn. Valenz .50



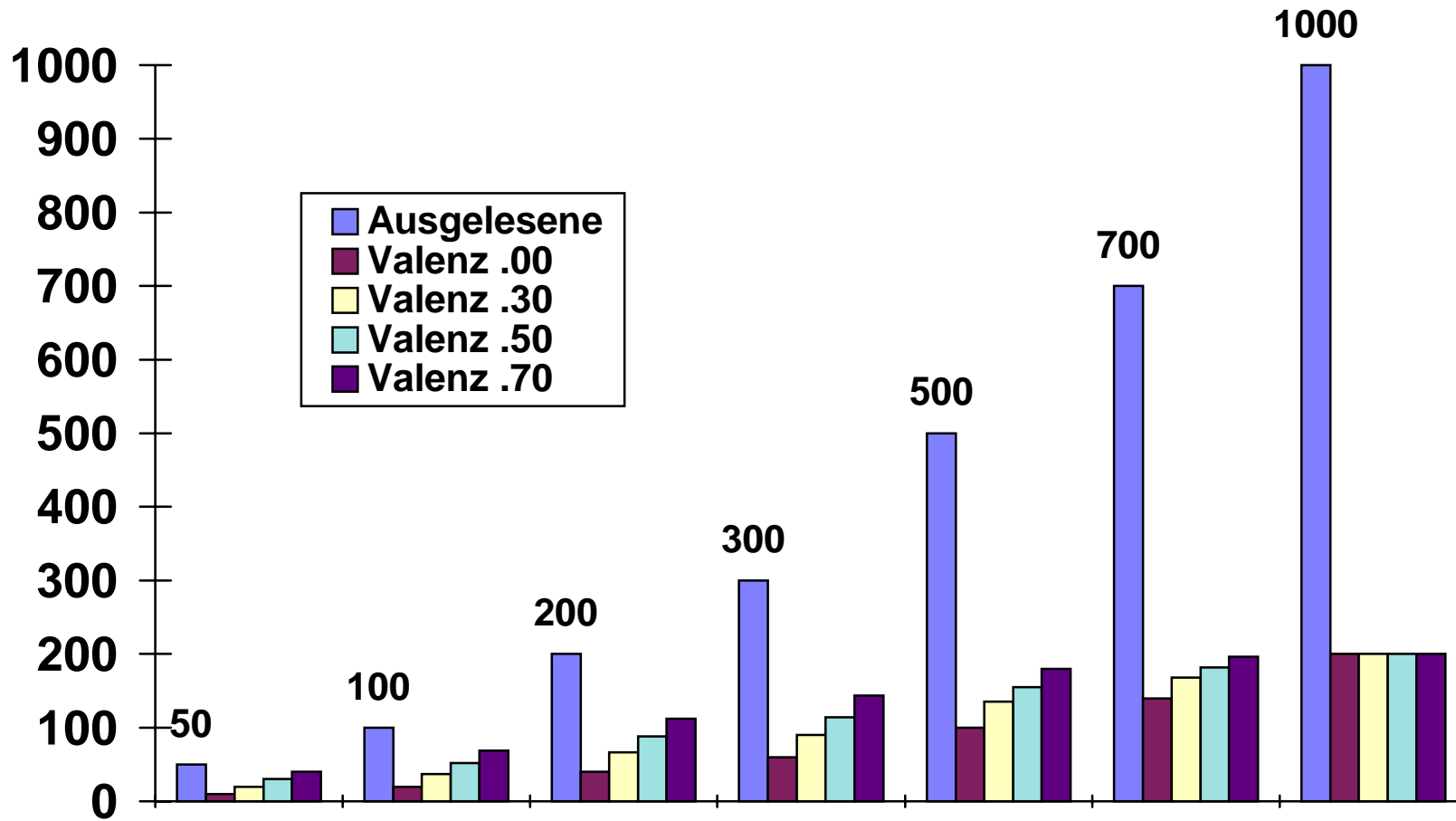
Nicht Geeignete, aber Eingestellte

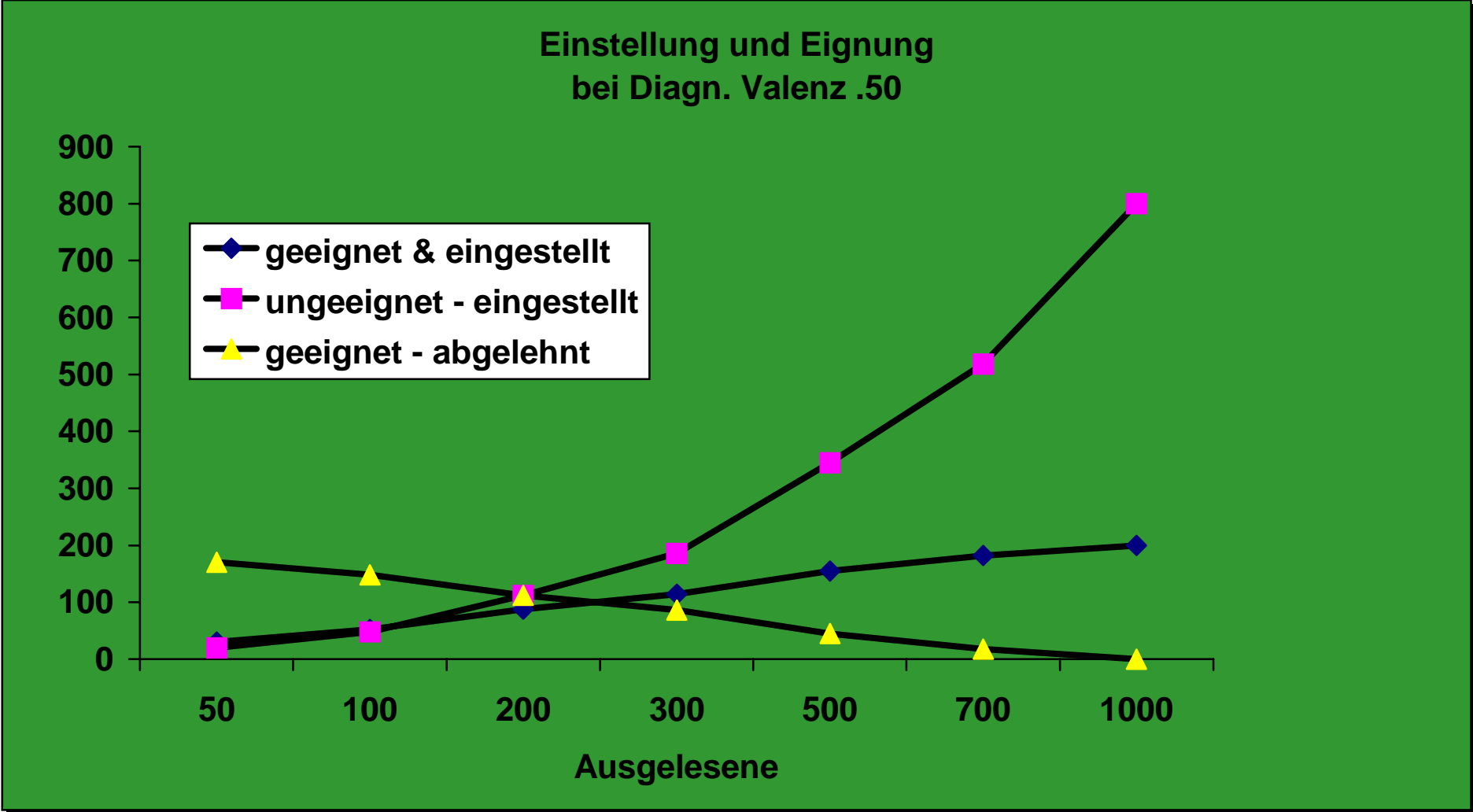


Nicht Eingestellte, aber Geeignete



Eingestellte Geeignete





Diagnostische Valenz		0,00			0,30			0,50			0,70		
		GE	GN	UE	GE	GN	UE	GE	GN	UE	GE	GN	UE
Ausgeseene	50	10	190	40	20	180	30	30	170	20	40	160	10
	100	20	180	80	37	163	63	52	148	48	69	131	31
	200	40	160	160	66	134	134	88	112	112	112	88	88
	300	60	140	240	90	110	210	114	86	186	144	56	156
	500	100	100	400	135	65	365	155	45	345	180	20	320
	700	140	60	560	168	32	532	182	18	518	196	4	504
	1000	200	0	800	200	0	800	200	0	800	200	0	800

¹⁾ Geeignete = G; Ungeeignete = U

²⁾ Eingestellte = E; Nicht Eingestellte = N

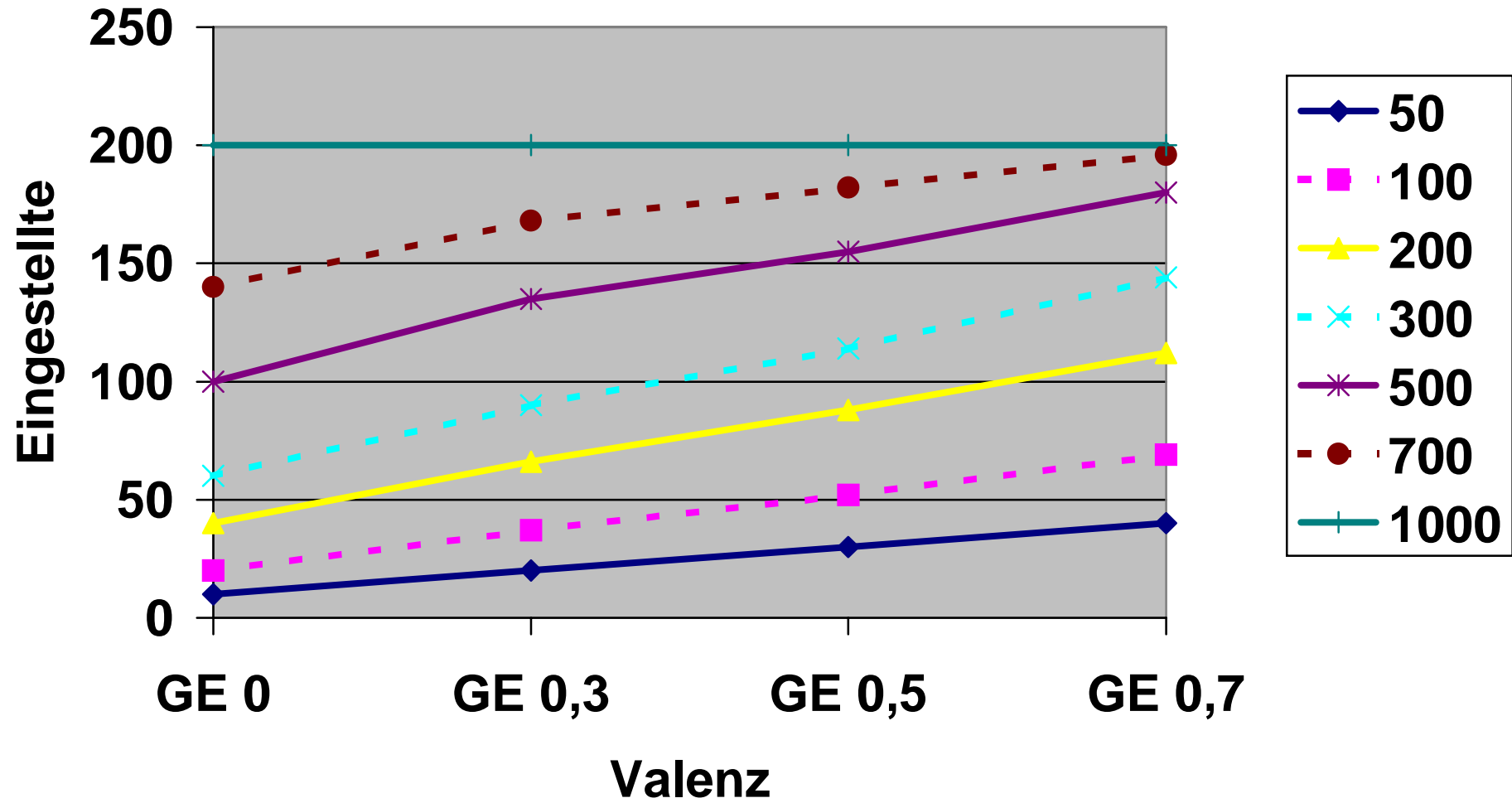
Diagnostische Valenz		0		0,3		0,5		0,7	
		GE	UE	GE	UE	GE	UE	GE	UE
Ausgeseene	50	10	40	20	30	30	20	40	10
	100	20	80	37	63	52	48	69	31
	200	40	160	66	134	88	112	112	88
	300	60	240	90	210	114	186	144	156
	500	100	400	135	365	155	345	180	320
	700	140	560	168	532	182	518	196	504
	1000	200	800	200	800	200	800	200	800

Auswahl	GE 0	UE 0	GE 0,3	UE 0,3	GE 0,5	UE 0,5	GE 0,7	UE 0,7
50	10	40	20	30	30	20	40	10
100	20	80	37	63	52	48	69	31
200	40	160	66	134	88	112	112	88
300	60	240	90	210	114	186	144	156
500	100	400	135	365	155	345	180	320
700	140	560	168	532	182	518	196	504
1000	200	800	200	800	200	800	200	800

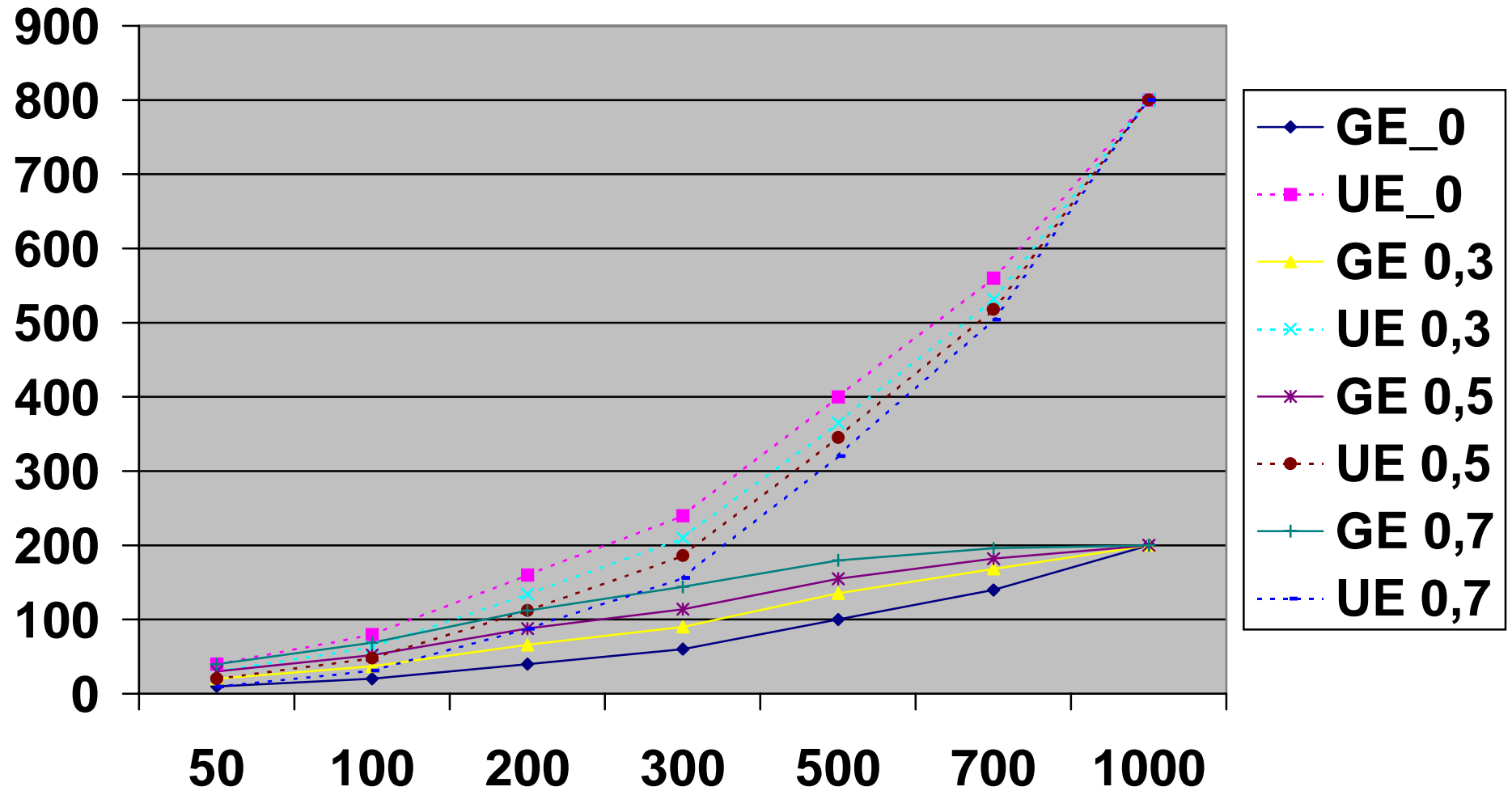
Auswahl	GE 0	GE 0,3	GE 0,5	GE 0,7
50	10	20	30	40
100	20	37	52	69
200	40	66	88	112
300	60	90	114	144
500	100	135	155	180
700	140	168	182	196
1000	200	200	200	200

ANZAHL DER EINGESTELLTEN GEEIGNETEN IN ABHÄNGIGKEIT VON VALENZ UND AUSWAHLGRÖÖBE:

Die testtechnisch sehr aufwendige Erhöhung der Valenz hat offensichtlich nur einen recht bescheidenen Erfolg im Vergleich zur blinden Anhebung der Auswahlgröße



Je größer die Gießkanne, desto nutzloser die Valenz

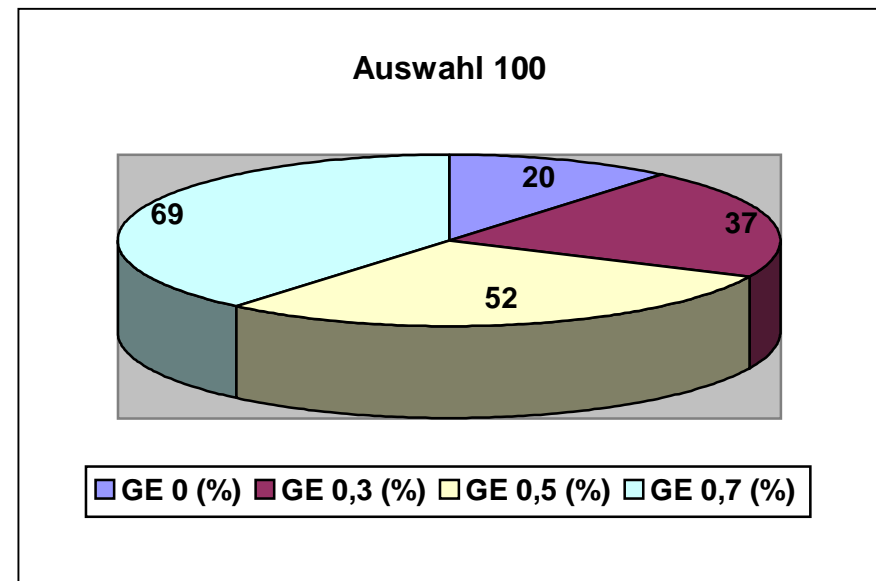
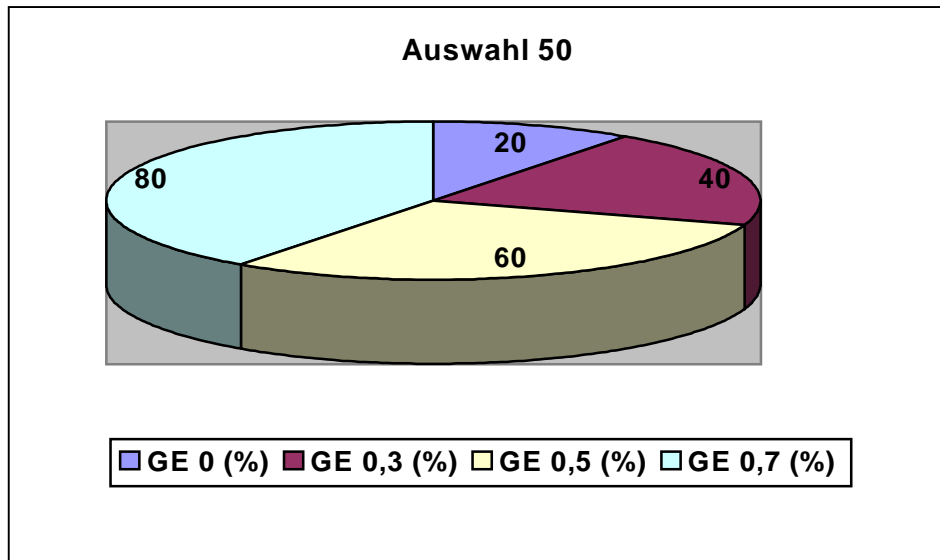


Auswahl	GE 0 (%)	GE 0,3 (%)	GE 0,5 (%)	GE 0,7 (%)
50	20	40	60	80
100	20	37	52	69
200	20	33	44	56
300	20	30	38	48
500	20	27	31	36
700	20	24	26	28
1000	20	20	20	20

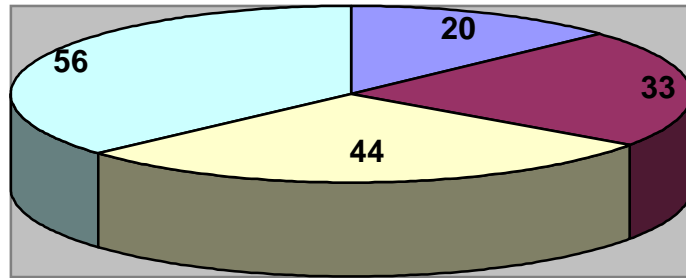
ANZAHL DER EINGESTELLTEN GEEIGNETEN IN ABHÄNGIGKEIT VON VALENZ UND AUSWAHLGRÖÖE:

Die Darstellung in Prozent zeigt deutlich, daß die Erhöhung der Valenz besonders bei kleinen Auswahlgrößen zum Tragen kommt.

Bei der Berechnung der Prozentzahlen wurde die Zahl der Ausgewählten gleich 100 % gesetzt.

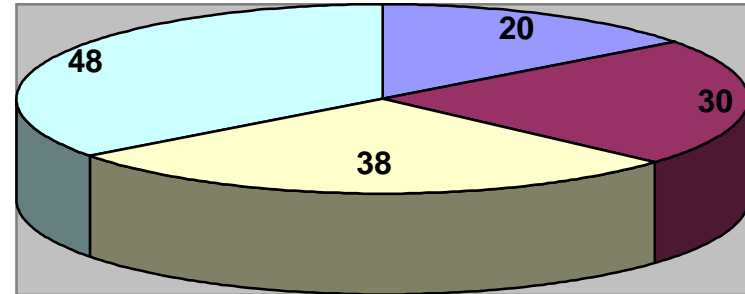


Auswahl 200



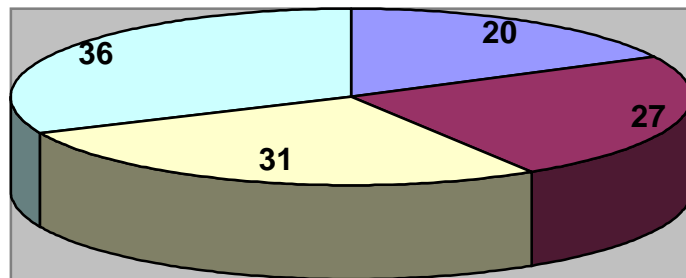
■ GE 0 (%) ■ GE 0,3 (%) ■ GE 0,5 (%) ■ GE 0,7 (%)

Auswahl 300



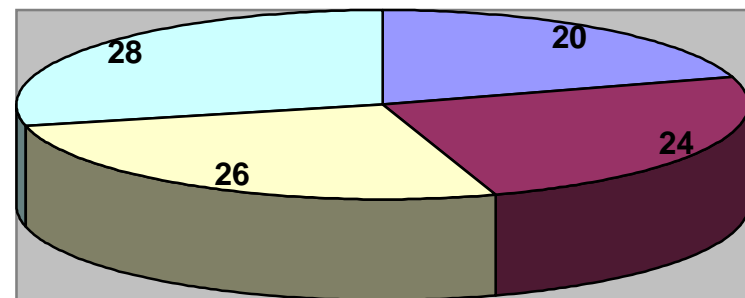
■ GE 0 (%) ■ GE 0,3 (%) ■ GE 0,5 (%) ■ GE 0,7 (%)

Auswahl 500

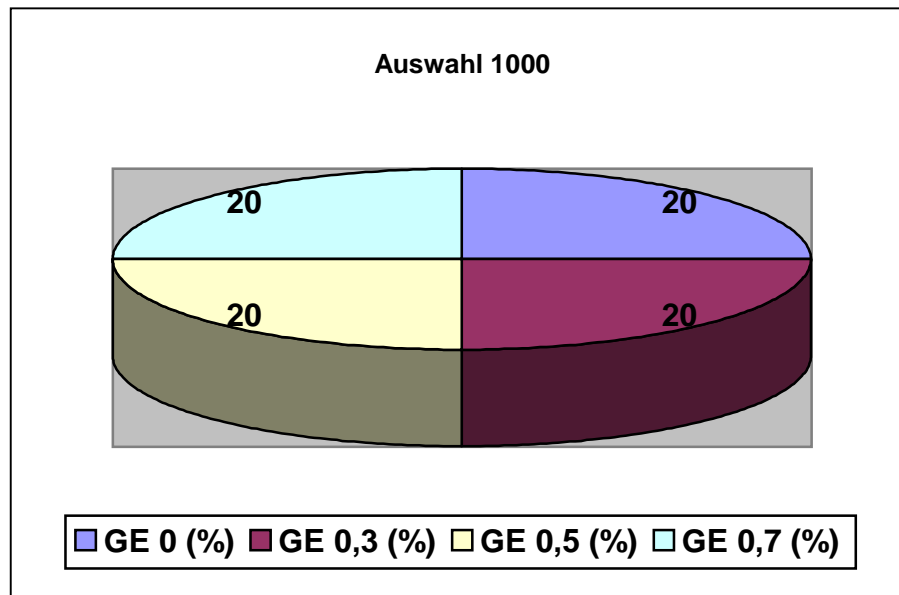


■ GE 0 (%) ■ GE 0,3 (%) ■ GE 0,5 (%) ■ GE 0,7 (%)

Auswahl 700



■ GE 0 (%) ■ GE 0,3 (%) ■ GE 0,5 (%) ■ GE 0,7 (%)



Die prozentuale Gleichberechtigung in jedem Fall hat ihren Preis, nämlich die Einstellung von immer viermal mehr Ungeeigneten — und die treiben in einer leistungsorientierten Gesellschaft jeden Betrieb ganz schnell in die Pleite. Und damit sind dann 100 % ohne Arbeit. Ganz ohne Statistik. Ganz sicher.

Viel problematischer als utopische Radikalprogramme sind für einen hauptamtlichen Diagnostiker wie z.B. einen leistungsmessenden Lehrer mit hoffentlich routiniertem und realistischem Pragmatismus die in der Ausgangstabelle erkenntlichen starken Einschränkungen der Verlässlichkeit auch von guten Testverfahren... Bei einer handwerklich sauberen Klassenarbeit kommt die gemessene Reliabilität übrigens selten über $r_{xy} = 0,5$ hinaus — Und bekanntlich kann die Validität eines Tests nie höher sein als seine Reliabilität.

Selbst bei einer für den Schulalltag traumhaften Valenz von 0,7 finde ich mit einem Test zur Ermittlung des Klassenbesten (1 von 20 analog zu 50 von 1000 in der Tabelle) zwar mit viermal höherer Wahrscheinlichkeit einen der Geeigneten als einen Ungeeigneten (40 vs. 10 in der ersten Zeile), aber ich lasse auch gleichzeitig vier ebenfalls Geeignete außen vor (40 vs. 160). Allerdings ist meine Trefferquote immerhin 16mal besser als bei einer Zufallsauswahl (40 auf 10 vs. 10 auf 40).

Und ob subjektive Auswahlen besser oder schlechter sind als zufällige, läßt sich nur über Tests entscheiden...