

Diagnose- und Vergleichsarbeiten 2007

Vergleich von Pilotierung 2006 und Pflichteinsatz 2007 in Bezug auf die prognostizierten und tatsächlichen Lösungshäufigkeiten

Die landesweiten Vergleichswerte der DVA 2007 und somit auch die Eckdaten für die Benotung basieren auf den Pilotierungsdaten 2006.

Aus den Erfahrungen des Vorjahres war bekannt, dass aufgrund der unterschiedlichen Rahmenbedingungen von Pilotierung und Pflichteinsatz mit fach- und schulartspezifisch unterschiedlich hohen Abweichungen zu rechnen ist. Diese Differenzen in den Lösungshäufigkeiten des Vorjahres (Pilotierung 2006 zu Pflichteinsatz 2006) wurden für jeden Test berechnet und bei der Ausweisung der Lösungshäufigkeiten aus der Pilotierung 2006 für die DVA 2007 berücksichtigt. Die Abweichungen in den Nicht-Kernfächern der Realschule und des Gymnasiums (vgl. Abb. 2-4, folgende Seiten) konnten vorab nicht abgeschätzt werden, da sie sich im Vorjahr auf Differenzen zwischen „Expertenurteil“ (Einschätzungen der Schwierigkeit durch die Mitglieder der Aufgabenfindungsgruppen ohne objektive Messung) und Pflichteinsatz bezogen haben. Es lagen somit keine empirisch gesicherten Erkenntnisse zu den Differenzen der Lösungshäufigkeiten in der Pilotierung und im Pflichteinsatz vor, um für die DVA 2007 zuverlässige Hinweise zur Anpassung der Vergleichswerte ableiten zu können.

Auch für die DVA 2007 wurden die Pilotierungsschulen gebeten, dem Landesinstitut Daten aus dem Pflichteinsatz zur Verfügung zu stellen, um so prognostizierte und tatsächliche Lösungshäufigkeiten miteinander vergleichen zu können.

In den nachfolgenden Abbildungen ist für die Vergleichsarbeiten in der Hauptschule, der Realschule und dem Gymnasium dargestellt, wie groß im jeweiligen Test der Anteil von Items (Teilaufgaben) mit mehr oder weniger großen Differenzen in den Lösungshäufigkeiten ist. Dabei werden 5 Differenzkategorien unterschieden:

- prozentualer Anteil der Items, die im Pflichteinsatz **viel schwerer** ausfallen („Ausreißer“), d.h. von deutlich weniger Schülerinnen und Schülern gelöst werden (Differenz größer -20%)
- prozentualer Anteil der Items, die im Pflichteinsatz **schwerer** ausfallen (Differenz zwischen -11% und -20%)
- prozentualer Anteil der Items, die im Pflichteinsatz der Pilotierung **vergleichbar** ausfallen (Differenz zwischen -10% und +10%)
- prozentualer Anteil der Items, die im Pflichteinsatz **leichter** ausfallen (Differenz zwischen +11% und 20%)
- prozentualer Anteil der Items, die im Pflichteinsatz **viel leichter** ausfallen („Ausreißer“), d.h. von deutlich mehr Schülerinnen und Schülern gelöst werden (Differenz größer +20%)



Landesinstitut
für Schulentwicklung
Rotebühlstraße 131
70197 Stuttgart
Fon 0711 6642-0
www.lis-bw.de

Qualitätsentwicklung
und Evaluation

Schulentwicklung
und empirische
Bildungsforschung

Bildungspläne

Abbildung 1: Hauptschule

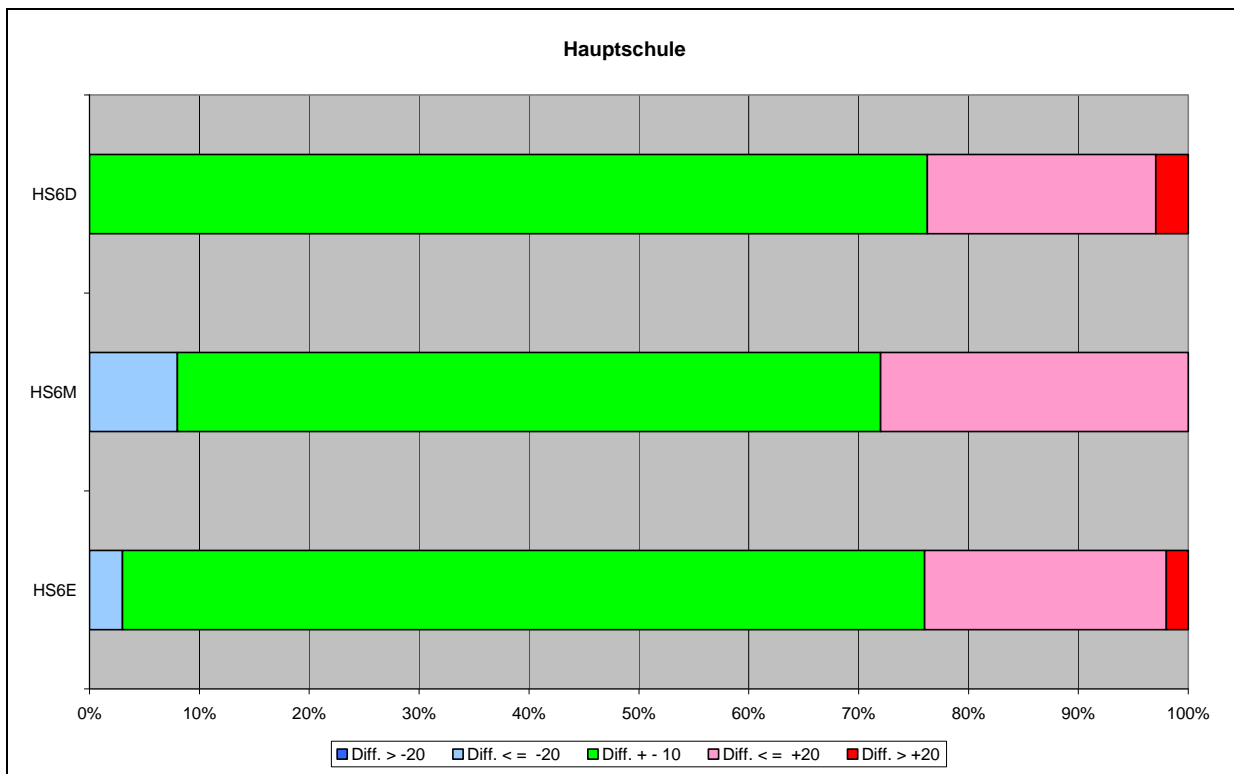
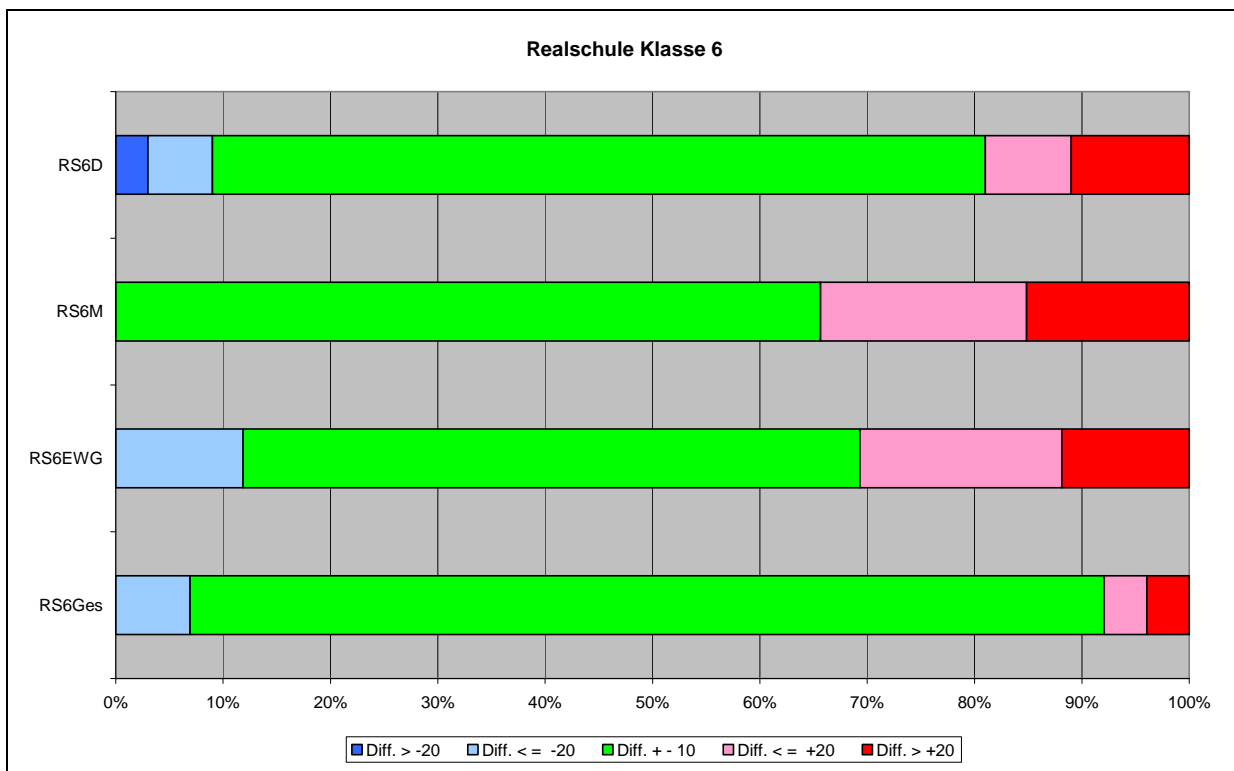
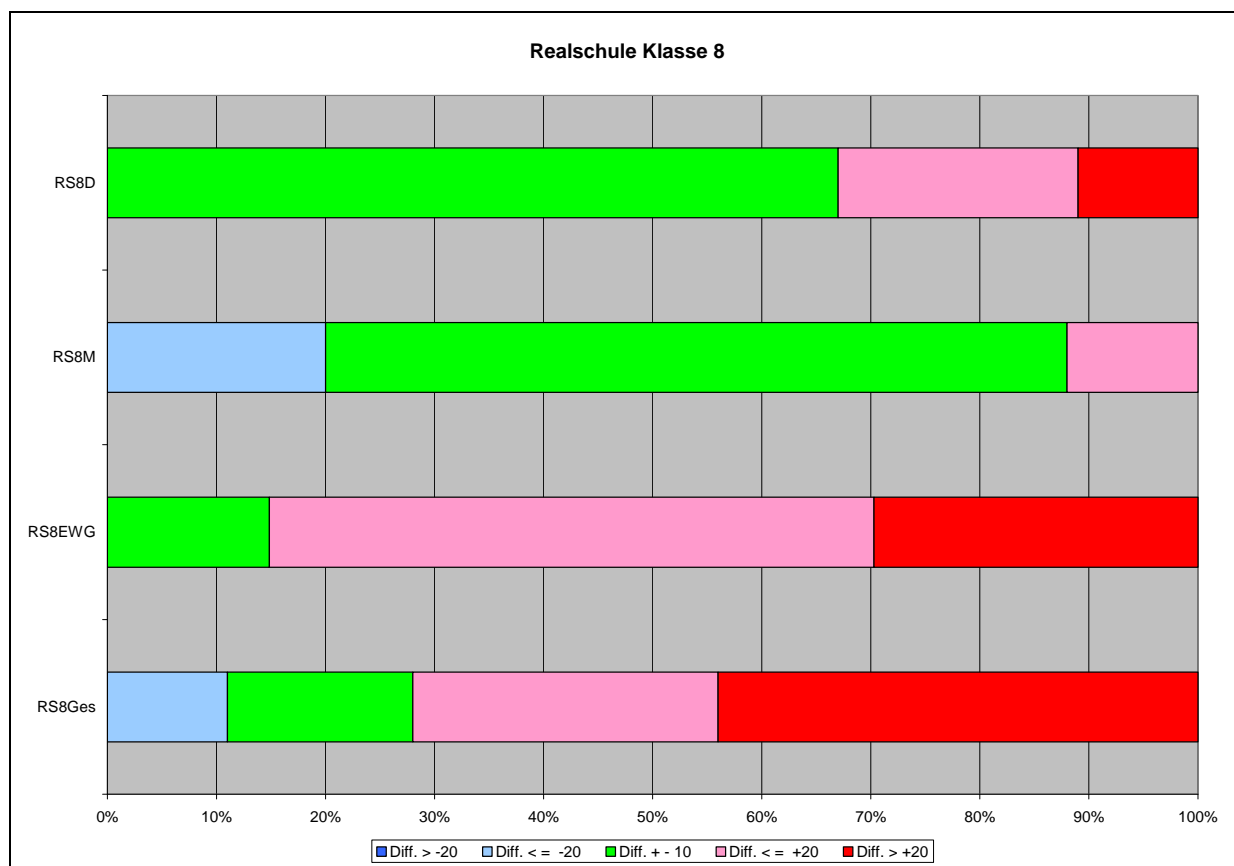


Abbildung 2: Realschule, Klasse 6



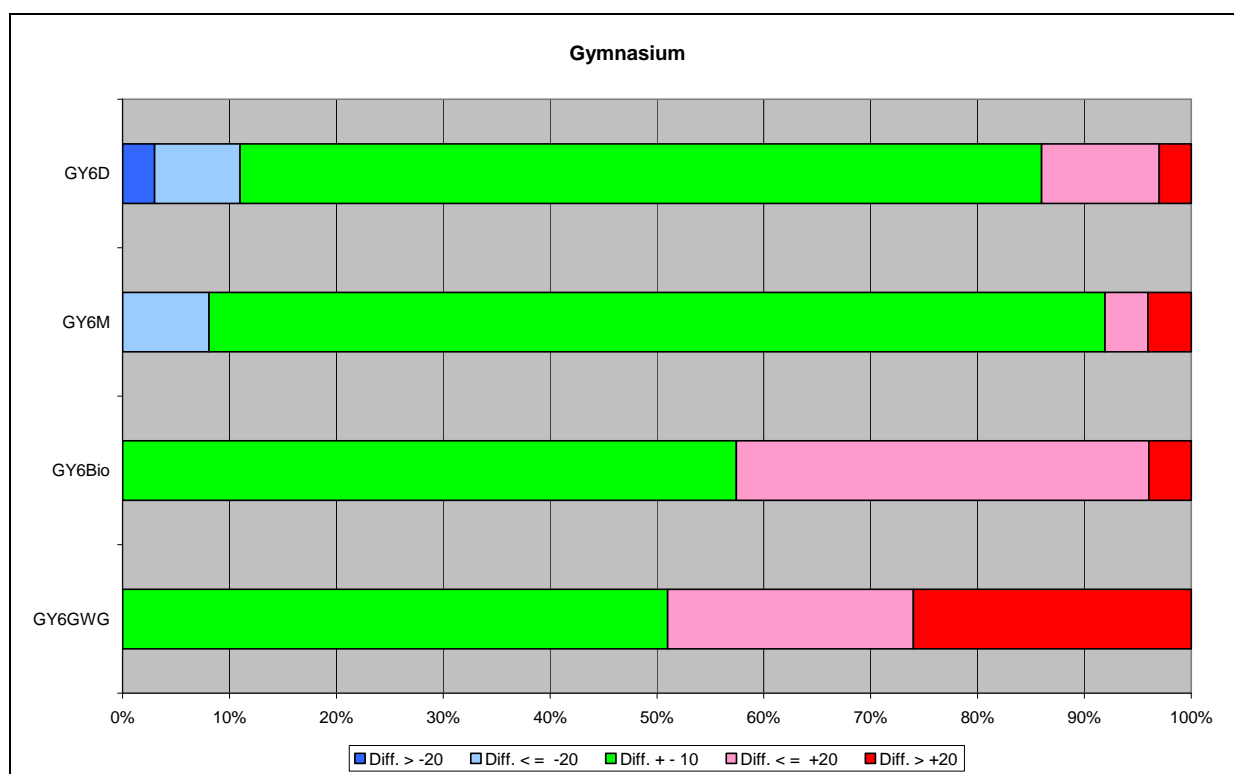
RS6EWG und RS6Ges in 2005 nicht pilotiert (Expertenurteil)

Abbildung 3: Realschule, Klasse 8



RS8EWG und RS8Ges in 2005 nicht pilotiert (Expertenurteil)

Abbildung 4: Gymnasium



GY6Bio und GY6GWG in 2005 nicht pilotiert (Expertenurteil)

Soweit für jedes Item die Differenz zwischen der für den Pflichteinsatz 2007 prognostizierten Lösungshäufigkeit und der tatsächlichen Lösungshäufigkeit im „grünen Bereich“ liegt, können die vorgenommenen Anpassungen als gelungen betrachtet werden. Tatsächlich ist der Anteil der Items innerhalb des grünen Bereichs bei den meisten Tests befriedigend hoch.

Für die Abweichungen in den Nicht-Kernfächern der Realschule und des Gymnasiums gilt dies nur eingeschränkt, da für diese Arbeiten aus 2005 nur „Expertenurteile“ vorlagen, eine empirisch fundierte Anpassung daher nicht möglich war.

In der nachfolgenden Tabelle werden für jeden Test die aufgrund der Pilotierung erwarteten durchschnittlichen Lösungshäufigkeiten (gemittelte Anzahl der gelösten Items) den im Pflichteinsatz tatsächlich beobachteten Lösungshäufigkeiten gegenübergestellt. Außerdem wird angegeben, welche Durchschnittsnote im Pflichteinsatz erreicht wurde.

Wieder sind es die Tests, für die aus dem Vorjahr keine empirischen Erkenntnisse zur Anpassung der Vergleichswerte vorlagen, die relativ große Differenzen bei den durchschnittlichen Lösungshäufigkeiten und den im Test erreichten Durchschnittsnoten aufweisen. Hinzu kommt der Deutschtest in Klasse 8 der Realschule, der im Pflichteinsatz deutlich besser ausgefallen ist als in der Pilotierung. Offensichtlich haben hier die vorgenommenen Anpassungen zu kurz gegriffen.

Tabelle 1: Gemittelte Anzahl der gelösten Items und Durchschnittsnote

Test	Anzahl Items	Lösung P2006	Lösung DVA2007	Note DVA2007 [#]
HS6D	39	27,2	29,4	2,8
HS6M	25	12,4	13,2	3,0
HS6E	60	30,6	33,4	3,0
RS6D	36	20,7	21,3	3,1
RS6M	26	9,9	11,7	2,8
RS6EWG *	26	12,1	13,2	2,9
RS6Ges *	28	14,2	14,5	3,1
RS8D	27	14,4	16,9	2,6
RS8M	25	10,1	10,0	3,2
RS8EWG *	27	10,2	14,6	2,3
RS8Ges *	18	7,8	10,5	2,4
GY6D	36	24,2	24,4	3,1
GY6M	24	12,8	12,5	3,2
GY6Bio *	26	13,7	15,8	2,7
GY6GWG *	39	22,5	27,8	2,4

* in 2005 nicht pilotiert (Expertenurteil), empirisch fundierte Anpassung daher nicht möglich

Vergleich mit Referenz-Note 3,0. Differenzen größer $\pm 0,2$ Notenpunkte sind rot gekennzeichnet.