

## 4 Zusammenfassung der Ergebnisse

### 4.1 International

Die vorliegende Untersuchung konnte den Zusammenhang von Testleistung und zwischen den Teilnehmerstaaten systematisch variierendem Schulalter mit einer Reihe statistischer Verfahren nachweisen. Die Auswahl der Verfahren orientierte sich dabei an den Vergleichsuntersuchungen TIMSS, PISA und IGLU. Die geringen Testwerte der Schülerinnen und Schüler im deutschen Sample sind als Ergebnis des Testdesigns, das heißt einer Schwerpunktsetzung zugunsten des Lebensalters zu bewerten.

Es hat sich gezeigt, dass weitere Variablen in der Auswertung berücksichtigt werden müssen: insbesondere das Schulalter, die unterschiedliche Rückhaltequote (Retentivität) der Schulsysteme, der Integrationsaspekt, die Bildungsexpansion, der Anteil zusätzlicher Beschulung (insbesondere in Japan und Korea) sowie die Verteilung der Variablen *Geschlecht*. Unter Berücksichtigung der Variablen *Klassenstufe* und *Geschlecht* erreichen bundesdeutsche Schülerinnen und Schüler Testleistungen im oberen Leistungsbereich der Vergleichsländer.

Die Studien TIMSS, PISA und IGLU weisen unterschiedliche Populationsdefinitionen auf, in der Regel eine Schwerpunktsetzung zugunsten einer Lebensalterdefinition, einer Schulalterdefinition oder einer Kombination aus beidem. Aus diesem Grund ist der direkte Vergleich der Testwerte einzelner Studien als längsschnittanalytischer Leistungsvergleich beziehungsweise als Messung von Leistungsfortschritten ungeeignet.

Der Technical Report verweist auf einen Mangel präziser Kenntnisse der zu untersuchenden Populationen. Hier wurde unter anderem Österreich angeführt. Vor diesem Hintergrund sind die Testwerte aber auch die Gewichtungszurteilung zu beurteilen.

Ein weiteres Problem stellt die variierende Rückhaltequote (Retentivität) der Schulsysteme dar. Ein positiver Zusammenhang zwischen Rückhaltequote (Retentivität) und Testleistung konnte nicht bestätigt werden. Besonders betroffen sind hier die Testleistungen Koreas.

Die Stichprobendefinition nach Lebensalter ist als wesentliche Ursache für *Disparitäten* anzusehen. Zudem sind so genannte Decken- und Bodeneffekte nachzuweisen. Es hat sich gezeigt, dass es bislang nicht gelungen ist, über eine Leistungsbandbreite von Schülern der 4. bis 13. Klassenstufe ein gleichermaßen geeignetes Testverfahren zu entwickeln. Die Schwerpunktsetzung zugunsten einer am Lebensalter orientierten

Stichprobendefinition angesichts erheblich unterschiedlich strukturierter Schulsysteme der Teilnehmerstaaten verstärkt die gemessene Leistungsheterogenität derart, dass Vergleiche nicht mehr zielführend erscheinen.

Die Definition der Stichprobe nach Lebensalter schafft für diejenigen Länder ohne strenge Koppelung von Lebens- und Schulalter, das sind mehr als zwei Drittel der Teilnehmerstaaten, Untersuchungspopulationen, die nicht mit schulischen Verhältnissen übereinstimmen. Man könnte unter Bezug auf schulische Verhältnisse von fiktiven Untersuchungspopulationen sprechen. Die Testergebnisse sind lediglich eine Folge der unterschiedlichen Gewichtung einzelner Schulalteranteile. Wenn diese über die verschiedenen Studien (PISA2000, PISA2003) relativ konstant bleiben, ändern sich auch die Testwerte nicht. Diese Konstanz als Reliabilität zu interpretieren ist nicht zielführend.

#### 4.2 National (Bundesrepublik)

Die differierenden mittleren Testergebnisse der einzelnen Bundesländer erweisen sich im wesentlichen als Ergebnis unterschiedlicher Bildungsexpansion. Es hat sich gezeigt, dass die berichteten Mittelwerte mit fortschreitender Bildungsexpansion negativ korreliert sind und sich daraus statistische Effekte ergeben (Stage Migration).

Der Vergleich der in PISA gezogenen Schülerpopulation mit Daten des Statistischen Bundesamtes weist Diskrepanzen sowohl in Bezug auf das Segment der Schülerinnen und Schüler mit Migrationshintergrund als auch in Bezug auf das Segment der Schüler im Dualen System der Berufsschulen auf. Es ist mit hoher Wahrscheinlichkeit davon auszugehen, dass in der Stichprobe derjenigen Bundesländer mit einem Anteil 15-jähriger Schüler in Berufsschulen, diese Gruppe nicht adäquat repräsentiert ist. Hieraus ergeben sich einerseits zum internationalen Vergleich gegensätzliche Zusammenhänge (Beispiel *Nichtleser*), andererseits überschätzte mittlere Testwerte. Ein direkter Vergleich dieser mittleren Werte ist daher ebenfalls nicht sinnvoll.

Daten des Statistischen Bundesamtes bezüglich der strukturellen Zusammensetzung der Bevölkerung zeigen einen kontinuierlichen anteiligen Rückgang der Kernfamilie an der Gesamtbevölkerung in moderaten Grenzen. Dieser Sachverhalt ist über die gesamten Teilnehmerstaaten auch querschnittsanalytisch nachzuweisen. Aus dieser mit steigendem sozioökonomischem Niveau kovariierenden soziodemographischen Entwicklung werden Konsequenzen bezüglich der Schülerleistungen zu erwarten sein.