

INDEX

- American Educational Research Association, 10
- American Psychological Association, 10
- benchmarking, 7–8
- bias
 - defined, 11–12
 - measurement, 11–15, 27
- Black students, performance of, 12–13, 34–35, 148, 156
- Chinese Taipei, 160–164. *see also* Taiwanese test takers and performance
 - differences in OTL patterns, 118–125, 142
 - educational policies, 172–173
 - omission patterns, 110–116
 - professional development activities for teachers, 174
 - test speededness, 116–117, 143
 - test-wiseness patterns, 108–118
- country selection, 25
 - country distinctions and the similarities, 62–74
 - criteria for selection, 38
 - distinct characteristics of countries, 66–67
 - educational history and reform efforts of countries, 39–40
 - factors impacting performance
 - analysis, 46
 - assessments, 60–62
 - measures, 44–45
 - school and instructional practices, 56–59
 - student-centered factors, 59–60
 - teacher factors, 51–56
 - test takers' home and background, 47–51
 - selected countries, 38–44
 - culture's impact on test-taking patterns, 33–35, 147–158, 166
 - analysis, 151–153
 - comparison of Black and White test takers, 148–149
 - high-context culture groups vs low-context culture groups, 152–157
 - measures, 151
 - research hypotheses, 150
- differential item function (DIF), 12–15, 24
 - comparison of Black and White test takers, 12–13
 - conditions to conduct, 14–15
 - definition, 12
 - explanations and sources of, 14

- educational accountability systems, 39
 8th grade assessment, 3, 5
 comparison of countries, 63–65,
 67–71, 73–78, 89–107
 cultural differences impact,
 151–152, 154, 156–158,
 163–164
 differences in OTL patterns,
 119–124
 high-context culture countries,
 158
 influencing factors
 home and background factors,
 47–51
 student-centered factors, 59–62
 teacher factors, 51–59
 item performance, 128, 132, 137,
 142, 144
 low-context culture countries, 157
 percentage of omissions, 111–
 116, 162
 performance patterns, 80–81,
 83–89
 test speededness, 116–118
 test-wiseness skills, 109–111
- fairness issues on assessments, xiv
*Filling in the Blanks: Understanding
 the Black White
 Achievement Gap*, xiii
- Finland, 160–164
 differences in OTL patterns,
 118–125, 142
 educational policies, 173
 omission patterns, 110–116, 145
 professional development
 activities for teachers, 175
 test speededness, 116–117, 143
 test-wiseness patterns, 108–118
 Finnish students, performance of
- application of mathematical
 procedures, 61–62
 average scale score and rank,
 87–88
 commonalities with other
 countries, 67–74
 distinct characteristics of
 performance, 66–67
 4th and 8th grade students,
 75–76
 frequency of assessments, 60–61
 influencing factors
 distractor choices, 144
 home and background factors,
 47–51
 item bias, 138–139
 nutrition issues, 50–51
 parental expectations of
 children's educational
 goals, 50
 school and instructional
 practices, 57–59
 student-centered factors, 59–60
 teacher factors, 51–56
 math achievement scores, 85
 vs. Qatari students, 95–96,
 104–106
 vs. Taiwanese students, 89–90,
 97–100
 vs. United States, 94–95,
 103–104
 p-value, 83–84, 128–129, 152
 4th grade assessment, 4–5
 comparison of countries, 63–66,
 69–70, 72, 74–78, 89–107
 cultural differences impact, 151,
 153, 155–157, 164
 differences in OTL patterns,
 119–124
 influencing factors

- home and background factors, 47–51
 - student-centered factors, 59–62
 - teacher factors, 51–59
- item performance, 127, 132, 135–136, 143
- percentage of omissions, 111–116, 162
- performance patterns, 80–82, 84–89
- test speededness, 116–118
- test-wiseness skills, 109–111
- Graduate Record Examination (GRE), 18
- high-context culture countries, and
 - test performance, 34–35, 38, 68–69, 77, 152–157, 166
- international assessments
 - benchmarking process, 7–8
 - comparison of major, 4
 - concerns and shortcoming of utilizing, 8–9
 - impact on education systems, 1–2 and policy, 6–7
 - test fairness, 9–10
- item bias, 11–12
 - identifying, 28–33, 128–129
 - sources of, 29–33, 130–135, 138–141
- item-level performance, 27–28
 - contextual issues, 29, 130, 141–142
 - differences in omission patterns, 144–145
 - distractor choices, 24, 31–32, 133–134, 144–145, 162–163, 166
 - identifying item bias, 14, 28–33, 128–129
 - measures, 127–128
 - OTL differences, 19–21, 26–27, 30, 118–126, 130–131, 142, 168, 170
 - sources of item bias, 29, 130–135, 138–141
 - strategy formation and usage, 28, 31, 132–133, 143–146, 162–163
 - test speededness, 26, 30–31, 116–117, 131–132, 142–143
 - test takers' omit patterns, 32–33, 110, 134–135, 144–145, 162–163
 - test wiseness, 108–118, 167–168
- low-context culture countries, and test performance, 34–35, 39, 68–69, 76–77, 152–157, 166
- mathematics test performance, 13, 31, 59–62, 168
 - of males and females in the United States, 13, 148
 - student-centered factors and, 59
 - variations in language and, 167
- National Council on Measurement in Education, 10
- No Child Left Behind Act (2002), 42–43
- opportunity to learn (OTL), 19–21, 26–27, 30, 118–126, 142, 168, 170
- parental expectations of children's educational goals, 47, 50, 68, 160

- PIRLS, 4, 48–50, 58
- predictive bias, 11–12
- Program for International Student Assessment (PISA), 2–3, 6–7
- impact of, 7
- PISA shock, 6
- report, use of, 7
- p*-value, or the proportion of test, 83–84, 128–129, 152
- Qatar, 160–164
- differences in OTL patterns, 118–125, 142
- educational policies, 173
- omission patterns, 110–116
- professional development
- activities for teachers, 174
- test speededness, 116–117, 143
- test-wiseness patterns, 108–118
- variations in language and impact on performance, 167
- Qatari students
- application of mathematical procedures, 61–62
- average scale score and rank, 87–88
- commonalities with other countries, 67–74
- distinct characteristics of performance, 66–67
- 4th and 8th grade students, 75–76
- frequency of assessments, 60–61
- impact of culture on test performance, 152–153
- influencing factors
- distractor choices, 144
- home and background factors, 47–51
- item bias, 137–138, 140
- language of the test, 49–50
- nutrition issues, 50–51
- own room and Internet access, 48
- parental expectations of children's educational goals, 50
- school and instructional practices, 57–59
- student-centered factors, 59–60
- teacher factors, 51–56
- math achievement scores, 86
- vs. Finnish students, 95–96, 104–106
- vs. Taiwanese students, 92–93, 101–103
- vs. United States, 96–97, 106–107
- p*-value, 83–84, 128–129, 152
- Science, Technology, Engineering, and Mathematics (STEM) fields and education, 3
- Standards for Educational and Psychological Testing, 10–21
- categorization of fairness, 11
- differential item function (DIF), 12–15
- general views of test fairness, 10–11
- item bias issues, 11–12
- purpose of, 10
- status of construct measured, 15
- student performance
- school and instructional practices, 56–59
- student-centered factors, 59–60
- teacher factors, 51–56
- test takers' home and background, 47–51

- Taiwanese test takers and
 performance, 40
 application of mathematical
 procedures, 61–62
 average scale score and rank,
 87–88
 commonalities with other
 countries, 67–74
 distinct characteristics of
 performance, 66–67
 4th and 8th grade students,
 74–76
 frequency of assessments, 60–61
 impact of culture on test
 performance, 152–153
 influencing factors
 distractor choices, 144
 home and background factors,
 47–51
 item bias, 136–140
 item context, 141–142
 language of the test, 49–50
 nutrition issues, 50–51
 own room and Internet access,
 48
 parental expectations of
 children's educational
 goals, 50
 teacher factors, 51–56
 math achievement scores, 85
 vs. Finnish students, 89–90,
 97–100
 vs. Qatari students, 92–93,
 101–103
 vs. United States, 91–92,
 100–101
 p-value, 83–84, 128–129, 152
 teaching to the test, 9
 test and subtest performance, 25–27,
 80–108
 average scale score for each
 country, 87
 for Chinese Taipei, 85, 87–93,
 97–103
 cycles of assessment
 administrations, 84–108
 definitions of content domains,
 81–83
 for Finland, 85, 87–91, 94–100,
 103–106
 magnitude of difference between
 countries, 83–84, 89–108
 p-value, or the proportion of
 test, 83–84
 mathematics test, 80–83
 measures, 80–83
 OTL differences, 26–27
 for Qatar, 86–88, 92–93, 95–97,
 101–107
 test wiseness, 26
 for United States, 86–88, 91–92,
 94–97, 100–101, 103–104,
 106–107
 test-based accountability measures,
 9, 38, 42–43, 78
 test developers, recommendations
 for, 170–171
 test fairness, issues in
 access to construct as
 measured, 15
 cultural impact, 33–35, 166
 equitable treatment, 11
 language, 49–50, 66–67, 70, 72,
 76, 167
 measurement bias, 11–15,
 165–166
 opportunity to learn, 19–21
 test content, 17
 test context, 17–18
 test response, 18–19

- test-wiseness skills, 167–168
- validity of interpretation, 15–19
- test speededness, 26, 30–31, 116–117, 142–143
- test wiseness skills, 26, 108–118, 167–168
- The Arbutnot Assessment Fairness (TAAF) Framework, 23–35. *see also* country selection; culture's impact on test-taking patterns; item-level performance; test and subtest performance
- phases, 23–35, 159–164
- recommendations to stakeholders, 169–175
 - for improving pedagogical practices and improve learning, 173–174
 - for policy making, 172–173
 - for test developers, 170–171
- strategy formation and usage, 31, 132–133, 143–145
- utilization of, 169
- 12th grade assessment, 3, 5
- Trends in International Mathematics and Science Study (TIMSS) assessment, 2–6
 - academic expectations, 60–61
 - content domains, 112–113
 - mathematics and science tests, domains of, 5
 - percentage of omissions, 110–111, 113–116
 - performance in different countries, 51–56, 89–107
 - questionnaire
 - academic expectations, 60
 - home and background factors, 45, 47–51
 - language skill, 48–49
 - learning environment, 59–62
 - learning mathematics, 59–60
 - own room and Internet access, 47–48
 - parental expectations of children's educational goals, 50
 - parent survey, 48–49
 - school and instructional practices, 56–59
 - teacher factors, 51–56
 - sampling strategy, 5
 - test speededness, 116–117
- United States, 160–164
 - differences in OTL patterns, 118–125, 142
 - educational policies, 172–173
 - impact of culture on test performance, 148–150, 156
 - omission patterns, 110–116
 - professional development
 - activities for teachers, 174
 - test speededness, 116–117, 143
 - test-wiseness patterns, 108–118
- United States students,
 - performance of
 - application of mathematical procedures, 61–62
 - average scale score and rank, 87–88
 - commonalities with other countries, 67–74
 - distinct characteristics of performance, 66–67
 - 4th and 8th grade students, 75–76
 - frequency of assessments, 60–61

- influencing factors
 - distractor choices, 144
 - home and background factors, 47–51
 - item bias, 136, 138–139
 - nutrition issues, 50–51
 - parental expectations of children’s educational goals, 50
 - school and instructional practices, 57–59
 - student-centered factors, 59–60
 - teacher factors, 51–56
- math achievement scores, 86
 - vs. Finnish students, 94–95, 103–104
 - vs. Qatari students, 96–97, 106–107
 - vs. Taiwanese students, 91–92, 100–101
- p*-value, 83–84, 128–129, 152
- White students, performance of, 12–13, 34–35, 148