



Support for Guidelines

Cara Cahalan Laitusis Educational Testing Service



Overview

- Background on how we decided on this format
- Types of documentation reviewed
- Review of support for each guideline
- Suggestions for
 - Additional guidelines to review
 - Research studies to support (or refute) each guideline
 - References that should be removed



Format

- NARAP Executive Committee discussed adapting “IES Practice Guidelines” Levels of Evidence
- Principles Committee suggested not providing levels of support
- Decision made by NARAP to use annotated bibliography format to provide support for each guideline but not to specify a level
- Support includes research studies as well as best practices from policy papers and other guidelines



Final Format

- Reference (alphabetical order)
- Summary paragraph of 150-175
 - Research Articles:
 - Purpose,
 - Method,
 - Results/Conclusions
 - Policy/Guidelines:
 - Purpose,
 - Main Points/Conclusion



Best Practice Documents Reviewed (guidelines, policy papers)

- National Association of School Psychologists (NASP): *Position Statement on Using Large Scale Assessments for High Stakes Decisions*
- Reports and Papers from National Center on Educational Outcomes (NCEO):
- National Agenda: *Moving Forward on Achieving Educational Equality for Deaf and Hard of Hearing Students*
- AERA: *Position Statement on High Stakes Testing in PreK -12 Education*
- AERA/APA/NCME: *Standards for Educational and Psychological Testing*
- ETS: *Standards for Quality and Fairness 2002*
- CTB/McGraw-Hill: *Guidelines for Inclusive Test Administration*



Guidelines and Policy

- CCSSO: Assessing Special Education Students: Training District and State Personnel on Accommodations
- APA Joint Committee on Testing Practices: Code of Fair Testing Practices in Education
- National Reading Conference Brief: High Stakes Testing and Reading Assessment (Afflerbach, P.)
- International Reading Association: *High-stakes assessments in Reading*
- The National Task Force on Equity in Testing Deaf and Hard of Hearing Students: Test accommodations for deaf and hard of hearing students
- American Printing House for the Blind (APH) *Test access: Making tests accessible for students with visual impairments: A guide for test publishers, test developers, and state assessment personnel* (Allman, C. B., 2004)



Research Studies

- Systematic literature review has NOT been conducted
 - NARAP researcher most familiar with guideline developed list based on prior knowledge
- Research studies have NOT been evaluated to determine merit of including each study



Guideline 1-A: Understand and account for the range of student characteristics and experiences ...

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association.
- Heubert, J., & Hauser, R. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Johnstone, C.J., Bottsford-Miller, N., & Thompson, S.J. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners* (Technical Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



Guideline 1-A. (cont.)

- Koenig, J.A., & Bachman, L.F. (2004). *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments*. Washington, DC: National Academies Press.
- Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Thompson, S.J., Johnstone, C.J., & Thurlow, M.L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M.L., Johnstone, C., Thompson, S.J., & Case, B.J. (in press). Using universal design research and perspectives to increase the validity of scores on large-scale assessments. In R. Johnson & R. Mitchell (Eds.), Washington, DC: Gallaudet Research Press.



Guideline 1-B: Begin the development of reading assessment by following the principles of universal design.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Allan, J. M., Bulla, N., Goodman, S.A., & Henderson, B. (2003). *Test access: guidelines for computer-administered testing*. American Printing House for the Blind: Louisville, KY.
- Baranowski, R.A. (2006). Item editing and editorial review. In S.M. Downing and T.M. Haladyna (Eds.). *Handbook of test development* (pp. 349-358). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.



Guideline 1-B. (cont.)

- Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment*, 3.
- Downing, S.M., & Haladyna, T.M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, 10, 61-82.
- Haladyna, T.M., & Downing, S.M. (2004). Construct irrelevant variance in high stakes testing. *Educational Measurement: Issues and Practice*, 23, 17-27.
- Henderson, B. W. (2005). *New perspectives on assessment of students who are blind or visually impaired. EnVision*, 10 add page numbers.



Guideline 1-B. (cont.)

- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Johnstone, C.J., Thompson, S.J., Miller, N.A., & Thurlow, M.L. (2008). Universal design and multi-method approaches to item review. *Educational Measurement: Issues and Practice*, 27, 25-36.
- Ketterlin-Geller, L. (2005). Knowing what all students know: Procedures for developing universally designed assessments. *The Journal of Technology, Learning, and Assessment*, 4, add page numbers.
- Leighton, J. P. (2004). Avoiding misconception, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, 23, 6-15.



Guideline 1-B. (cont.)

- Linn, R.L., & Harnsisch, D.L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement*, 18, 109-118.
- Messick, S. (1996). *Standards-based score interpretation: Establishing valid grounds for valid inferences*. Proceedings of the Joint Conference on Standard Setting for Large Scale Assessments, Sponsored by National Assessment Governing Board and The National Center for Education Statistics. Washington, DC: Government Printing Office.
- Zieky, M. (2006). Fairness reviews in assessment. In S.M. Downing and T.M. Haladyna (Eds.). *Handbook of test development* (pp. 359-376). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.



Guideline 1-C. Use technologies and other evidence-based approaches to provide all students with a similar array of assessment options.

- Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice*, 16, 174-181.
- Minnema, J., Thurlow, M., & Hopfengardner Warren, S. (2004). Understanding out-of-level testing in local schools: A first case study of policy implementation and effects. NCEO Out-of-Level Testing Project Report 11.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75, 457-490.
- Zenisky, A. L., & Sireci, S. G. (2007). A summary of the research on the effects of test accommodations: 2005-2006. (Technical Report 47). Minneapolis, MN: National Center on Educational Outcomes.



Guideline 1-D. Document decisions that are made in the name of accessibility, and monitor their effects for different group of students.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association.
- National Academies Press. (2007). *Lessons learned about testing: Ten years of work at the National Research Council*. Washington, DC: Author.



Guideline 2-A. State standards should be grounded in an evidence-based definition of reading.

- Afflerbach, P. (2004). *National Reading Conference Policy Brief: High Stakes Testing and Reading Assessment*. Oak Creek, WI: The National Reading Conference.
- International Reading Association. *High Stakes Assessments in Reading: A Position Statement of the International Reading Association* (1999). Newark, DE: Author
- McKenna, M.C., & Stahl, S. A. (2003). *Assessment for reading instruction*. New York: Guilford Press.
- Moje, E. B., Dillon, D. R., & O'Brien, D. G. (2000). Re-examining roles of learner, text, and context in secondary literacy. *Journal of Educational Research*, 93, 165-180.



Guideline 2-A. (cont.)

- Pearson, P. D. (1984). *Handbook of reading research*. New York: Longman.
- Rummelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed), *Attention and performance* (Vol. 6, pp. 573-603). Hillsdale, NJ: Erlbaum.
- Stahl, S. A., Kuhn, M. R., & Pickle, M. J. (1999). An educational model of assessment and targeted instruction for children with reading problems. In D. Evenson & P. Mosenthal (Eds.), *Reconsidering the role of the reading clinic in a new age of literacy* (pp. 249-272). Stamford Press, CT: JAI Press.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford.



Guideline 2-B. Ensure that test blueprints are aligned with the state standards.

- Johnstone, C.J., Thurlow, M.L., Thompson, S.J., & Clapper, A.T. (2008). The potential for multi-modal approaches to reading for students with disabilities as found in state reading standards. *Journal of Disability Policy Studies, 18* , 219-229.
- Johnstone, C. J., Moen, R. E., Thurlow, M. L., Matchett, D., Hausmann, K. E., & Scullin, S. (2007). *What do state reading test specifications specify?* Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment.
- La Marca, P. M. (2001). Alignment of standards and assessments as an accountability criterion. *Practical Assessment, Research & Evaluation, 7* add page numbers.
- National Assessment Governing Board. (2005). *Specifications for the 2009 NAEP reading assessment.* Washington, DC: Author.



Guideline 2-B. (cont.)

- Thurlow, M.L., Johnstone, C.J., Thompson, S.J., & Case, B. (in press). Using universal design research and perspectives to increase the validity of scores on large-scale assessments. In Karchmer, M. (Ed.). *Assessing deaf students' academic achievement in an age of accountability*, add page numbers. Washington, DC: Gallaudet University Press.
- R. Johnson & R. Mitchell (Eds.), *Assessing deaf students' academic achievement in an age of accountability*. Washington, DC: Gallaudet Research Press.
- Webb, N.L. (1999). *Alignment of Science and Mathematics standards and assessments in four states*. Washington, DC: Chief Council of State School Officers.
- Webb, N.L. (2006). Identifying content for student achievement tests. In S.M. Downing and T.M. Haladyna (Eds.). *Handbook of test development* (pp. 155-180). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.



Guideline 2-C. Design reading tests to allow for individual measurement of the components reflected in state standards.

- Afflerbach, P. (2004): *High stakes testing and reading assessment*. (NRC Policy Brief), Oak Creek, WI: The National Reading Conference.
- Fletcher, J.M., Lyon, G.R., Fuchs, L.S., & Barnes, M.A. (2007). *Learning disabilities from identification to intervention*. New York: Guilford Press.
- Laitusis, C.C., & Cook. L.L. *Reading aloud as an accommodation for a test of reading comprehension. Research Spotlight, 1, 15-20*. Princeton, NJ: Educational Testing Service.
- PARA Reports see www.readingassessment.info/resources/index.htm
 - Reading and Students with Autism;
 - Reading and Students with Visual Impairments of Blindness;
 - Reading and Students who are Deaf or Hard of Hearing.



Guideline 2-D. Use multiple criteria to select texts and visual elements.

- Afflerbach, P. (2004). *High stakes testing and reading assessment*. (NRC Policy Brief) Oak Creek, WI: The National Reading Conference.
- Allman, C. (2004). *Making tests accessible for students with student with visual impairments: A guide for test publishers, test developers, and state assessment personnel*. (2nd edition). Louisville, KY: American Printing House for the Blind. Available from
- Chall, J. & Squire, J. R. (1996). The publishing industry and textbooks. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson. *Handbook of reading research* (Vol. II, pp. 120-146). New York: Longman.
- Donald, (1979). Effects of illustrations on early oral reading accuracy, strategies and comprehension. *British Journal of Educational Psychology*, 49, 282-289.



Guideline 2-D. (cont.)

- Goldman, S., & Rakestraw, J. (2000). Structural aspects of constructing meaning from text. In R. Barr, M. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. III, pp. 311–335). New York: Longman.
- Johnson, P. (1984). Assessment in reading. In P.D. Pearson, (Ed.), *Handbook of reading research* (pp. 147-182). New York: Longman.
- Levie, W. H., & Lentz, R. (1982). Effect of text illustrations: A review of research. *Educational Communication and Technology*, 30, 195-232.
- Willows, D. M., Borwick, D., & Hayren, M. (1981). The content of school readers. In G. E. Mackinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 2, pp. 100-179). New York: Academic Press.



Guideline 2-D. (cont.)

- Meyer, B. J. F., & Rice, C. E. (1984). The structure of text. In P.D. Pearson (Ed.), *Handbook of reading research* (pp. 319-351). New York: Longman.
- Armbruster, B. (1994). The problem of inconsiderate text. In G. Duffy, L. Roehler, & J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions* (p. 203). New York: Longman.
- Pearson, P.D., & Camperell, K. (1994). Comprehension of text structures. In R.B. Ruddell, M.R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes at reading* (4th ed., pp. 448–468). Newark, DE: International Reading Association.
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 545–586). Mahwah, NJ: Erlbaum.



Guideline 2-D. (cont.)

- Purves, A.C. (1973). Structural aspects of constructing meaning from text. In R. Barr, M. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. III, pp. 311–335). New York: Longman.
- Shallert, D. L., & Reed, J. H. (1997). The pull of text and the process of involvement in reading. In J. Guthrie & A. Wigfield (Eds.), *Reading engagement: Motivating readers through integrated instruction* (pp. 68-85). Newark, NJ: International Reading Association.
- Simons, H. D., & Elster, C. (1990). Picture dependence in first-grade basal text. *Journal of Educational Research*, 84, 86-92.
- Waller, R. (1996). Typology and discourse. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson. (Eds.), *Handbook of reading research*, (Vol. II, pp. 341-380). New York: Longman.



Guideline 2-E. Include interesting texts that are likely to engage all students.

- Hall, C., & Cole, M. (1999). *Children's reading choices*. London: Routledge.
- Moss, G., & McDonald, J. W. (2003). The borrowers: Library records as unobtrusive measures of children's reading preferences. *Journal of Research in Reading, 27*, 401-412.
- Mohr, K. (2006). Children's choices for recreational reading: A three-part investigation of selection preferences, rationales, and process. *Journal of Literacy Research, 38*, 81-104.
- National Assessment Governing Board. (2005). *Specifications for the 2009 NAEP reading assessment*. Washington, DC: Author. Retrieved February 18, 2006 from
- Salinger, T., Kamil, M. L., Kapinus, B., & Afflerbach, P. (2005). Development of a new framework for the NAEP reading assessment. *54th Yearbook of the national reading conference* (pp. 334-348). Oak Creek, WI: National Reading Conference.



Guideline 2-F. Present reading tasks that students perceive as worthwhile and that enable them to be self-efficacious as learners.

- Guthrie, J. T., Van Meter, P., Hancock, G., Alao, S., Anderson, E., & McCann, A. (1998). Does concept-oriented reading instruction increase strategy use and conceptual learning from text. *Journal of Educational Psychology, 90*, 261-278.
- Kincade, K. M. (1991). Patterns in children's ability to recall explicit, implicit, and metaphorical information. *Journal of Research in Reading, 14*, 81-98.
- Paris, S.G., & Carpenter, R. D. (2004). Children's motivation to read. In J. V. Hoffman & D. L. Schallert (Eds.), *The texts in elementary classrooms* (pp. 61-82). Mahwah, NJ: Erlbaum.
- Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing instruction. *Contemporary Educational Psychology, 18*, 337-354.
- Turner, J. C. (1995). The influence of classroom contexts on young children's motivation for literacy. *Reading Research Quarterly, 30*, 410-441.



Guideline 3-A. Consider all test takers' needs when developing the initial design for the test.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Schmeiser, C. B., & Welch, C. J. (2006). Test Development. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 307-353). Westport, CT: American Council on Education & Prager Publishing.



Guideline 3-B. Consider all test takers' needs in developing and evaluating assessment items and tasks.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Koenig, J.A., & Bachman, L.F. (2004). *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments*. Washington, DC: National Academies Press.
- Hansen, E. G., & Mislevy, R. J. (2007). Accessibility of computer-based testing for individuals with disabilities and English language learners within a validity framework. In L. Tomei, (Ed.) *Online and distance learning: Concepts, methodologies, tools, and applications* (pp. 2532-2567). Hershey, PA: Idea Group Publishing,



Guideline 3-B. (cont.)

- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Johnstone, C.J., Thompson, S.J., Miller, N.A., & Thurlow, M.L. (2008). Universal design and multi-method approaches to item review. *Educational Measurement: Issues and Practice*, 27, 25-36.
- Kane, M. T. (2006). Validation. In R. L. Brennan, *Educational measurement* (4th ed), (pp. 17-64). Westport, CT: Praeger.



Guideline 3-C. Consider all test takers' needs in assembling and evaluating tests.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Johnstone, C. J., Moen, R. E., Thurlow, M. L., Matchett, D., Hausmann, K. E., & Scullin, S. (2007). *What do state reading test specifications specify?* Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment.
- Johnstone, C.J., Thurlow, M.L., Thompson, S.J., & Clapper, A.T. (2008). The potential for multi-modal approaches to reading for students with disabilities as found in state reading standards. *Journal of Disability Policy Studies, 18* , 219-229.
- Webb, N.L. (1999). *Alignment of Science and Mathematics standards and assessments in four states*. Washington, DC: Chief Council of State School Officers.



Guideline 3-D. Document the steps that have been taken to ensure that all test takers' needs have been considered.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Allan, J. M., Bulla, N., Goodman, S.A., & Henderson., B. (2003). *Test access: guidelines for computer-administered testing*. American Printing House for the Blind: Louisville, KY.



Guideline 4-A. Begin the assessment development or revision process by reviewing allowed accommodations to determine whether they could be incorporated into the design of the assessment.

- Koenig, J.A. (Ed.), (2002). *Reporting test results for students with disabilities and English-language learners*. Washington, DC: National Research Council.
- Thompson, S.J., Johnstone, C.J., & Thurlow, M.L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M.L., Thompson, S.J., & Lazarus, S.S. (2005). Considerations for the administration of tests to special needs students: Accommodations, modifications, and more. In S.M. Downing & T.M. Haladyna (Eds), *Handbook of test development*. Add page numbers Mahwah, NJ: Lawrence Erlbaum.



Guideline 4-B. Identify and determine the essential accommodations that are still needed after incorporating as many as possible into the assessment.

- Thompson, S.J., Quenemoen, R.F., & Thurlow, M.L. (2006). Factors to consider in the design of inclusive online assessments. In M. Hricko & S.L. Howell (Eds.), *Online assessment and measurement: Foundations and challenges*. Hershey, PA: Information Science Publishing.



Guideline 4-C. Develop a strong rationale and evidence to support the validity of inferences from assessment results when accommodations are provided.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association.



Guideline 4-D. Provide information and support to ensure that accommodations are implemented in an appropriate manner.

- CTB McGraw-Hill. (2005). *Guidelines for inclusive test administration*. Monterey, CA, Author.



Guideline 4-E. Adjust the reading assessment approach to address the needs of some groups of students that can not be met by typical test design or accommodation procedures.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Allman, C. B. (2004). Test access: Making tests accessible for students with visual impairments: *A guide for test publishers, test developers, and state assessment personnel*. American Printing House for the Blind. Louisville, KY.
- CTB McGraw-Hill. (2005). *Guidelines for inclusive test administration*. Monterey, CA, Author.



Guideline 5-A. Provide clear and concise score reports that are appropriate for a diverse audience.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Aschbacher, P.R., & Herman, J.L. (1991). *Guidelines for effective score reporting* (CSE Tech. Rep. No. 326). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.
- Goodman, D.P., & Hambleton, R.K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17, 145-220.
- Hambleton, R.K., & Slater, S. (1997). *Are NAEP executive summary reports understandable to policy makers and educators?* (CSE Technical Report 430). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Teaching.



Guideline 5-A. (cont.)

- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association.
- Trout, D.L., & Hyde, E. (2006, April). *Developing score reports for statewide assessments that are valued and used: Feedback from K-12 stakeholders*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Wainer, H., Hambleton, R.K., Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement* 36, 301–335.



Guideline 5-B. Pilot score reports with all relevant groups of score users.

- Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17 , 145-220.



Guideline 5-C. Detailed information about the assessment and score results should be included in a separate guide or technical manual that is accessible to all test takers and score users.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Goodman, D.P., & Hambleton, R.K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17, 145-220.
- Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, & M. R., Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice* 10, 16–18.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association.



Guideline 5-D. Provide information regarding the precision of reported scores for all relevant groups.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association.
- Educational Testing Service. *ETS Standards for quality and fairness*. (2002). Princeton, NJ: Author.
- National Association of School Psychologists. *Large scale assessments and high stakes decisions: Facts, cautions and guidelines*.



- Final Comments?
- Suggestions on support document as a whole?