NATIONAL
ACADEMY
of
EDUCATION

WORKSHOP ON BIG DATA IN EDUCATION: BALANCING THE BENEFITS OF EDUCATIONAL RESEARCH AND STUDENT PRIVACY

COMPILATION OF REFERENCES AND RESOURCE LINKS

Aist, G., & Mostow, J. (2000). Using Automated Within-Subject Invisible Experiments to Test the Effectiveness of Automated Vocabulary Assistance. In *Proceedings of the ITS'2000 Workshop on Applying Machine Learning to ITS Design/Construction, Fifth International Conference on Intelligent Tutoring Systems*, pp. 4-8.

ASSISTmentsData. (n.d.). Datasets. Available at https://sites.google.com/site/assistmentsdata.

Baron, J., Figlio, D., Goldhaber, D., Haskins, R., Mays, A., Raymond, M., and Vance, A. (2016). Turning Data into Information: The Vital Role of Research in Improving Education. *Data Quality Campaign*. Available at http://2pido73em67o3eytaq1cp8au.wpengine.netdna-cdn.com/wp-content/uploads/2016/08/DQC-Why-research-matters-09142016.pdf.

Beck, J. E., & Mostow, J. (2008). How Who Should Practice: Using Learning Decomposition to Evaluate the Efficacy of Different Types of Practice for Different Types of Students. In *Intelligent Tutoring Systems*, Lecture Notes in Computer Science Vol. 5091. New York: Springer. Pp. 353-362.

Boston College. (n.d.). edX Evaluation Study. Available at http://www.bc.edu/research/csteep/edx.html.

Brundin, J. (2015). Colorado Student Survey on Drugs, Sex, Other Health Issues Causes Stir. *Colorado Public Radio.* Available at http://www.cpr.org/news/story/colorado-student-survey-drugs-sex-other-health-issues-causes-stir.

Chabris, C. F., & Meyer, M. N. (2016, June 19). Please, Corporations, Experiment on Us. *The New York Times*. Available at http://www.nytimes.com/2015/06/21/opinion/sunday/please-corporations-experiment-on-us.html?_r=0.

Chandler, D. L. (2014). Study: Online Classes Really Do Work: Detailed Quantitative Study of Learning Online Shows Success, Even Among Those with Least Preparation. *MIT News*. Available at http://news.mit.edu/2014/study-shows-online-courses-effective-0924.

Chingos, M. M. (2016). No More Free Lunch for Education Policymakers and Research-ers. *Evidence Speaks Reports* 1(20). Available at https://www.brookings.edu/wp-content/uploads/2016/06/free-and-reduced-lunch3.pdf.

Commission on Evidence Based Policymaking. (n.d.). Available at https://www.whitehouse. gov/omb/management/commission_evidence.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction* 4(4):253-278.

Cornell University Law Center Legal Information Institute. (n.d.). 20 U.S.C. § 1232g. Family Educational and Privacy Rights. Available at https://www.law.cornell.edu/uscode/ text/20/1232g.

Data Quality Campaign. (2016). Student Data Privacy Legislation: A Summary of 2016 State Legislation. Available at http://2pido73em67o3eytaq1cp8au.wpengine.netdnacdn.com/ wp-content/uploads/2016/09/DQC-Legislativesummary-09302016.pdf.

Data Quality Campaign. (2016). Turning Data into Information: The Vital Role of Education Research in Improving Education. Available at http://2pido73em67o3eytaq1cp8au. wpengine.netdnacdn.com/wp-content/uploads/2016/08/DQC-Why-research matters-09142016.pdf.

Databrary. Available at https://nyu.databrary.org.

Davis, E., & Marcus, G. (2014). Eight (No, Nine!) Problems with Big Data. *The New York Times*. Available at http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html.

Dede, C. (Ed.). (2015). Data-Intensive Research in Education: Current Work and Next Steps: Report on Two National Science Foundation-Sponsored Computing Research Education Workshops. Washington, DC: Computing Research Association. Available at http://cra. org/wp-content/uploads/2015/10/CRAEducationReport2015.pdf.

Effrem, K. R., & Robbins, J. (2015). Response to Concerns About the Student Privacy Protection Act—S. 1341. *Education Liberty Watch*. Available at http://edlibertywatch.org/2015/06/ response-to-concerns-about-the-student-privacy-protection-act-s-1341.

Electronic Privacy Information Center. (n.d.). Student Privacy Bill of Rights. Available at https://epic.org/privacy/student/bill-of-rights.html.

Executive Office of the President. (2014). Big Data: Seizing Opportunities, Preserving Values. Washington, DC. Available at https://obamawhitehouse.archives.gov/sites/default/ files/docs/big_data_privacy_report_may_1_2014.pdf.

Family Policy Compliance Office. (2016). Protection of Pupil Rights Amendment (PPRA). Available at http://familypolicy.ed.gov/ppra.

Federal Policy for the Protection of Human Subjects ("Common Rule"). Available at https:// www.hhs.gov/ohrp/regulations-andpolicy/regulations/common-rule/index.html.

Federal Trade Commission. (2016). Complying with COPPA: Frequently Asked Questions. Available at https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequentlyasked-questions.

Feng, M., Heffernan, N., & Koedinger, K. R. (2009). Addressing the Assessment Challenge with an Online System That Tutors as It Assesses. *User Modeling and User-Adapted Interaction* 19(3)243-266.

Frank, L. (2000). When an Entire Country Is a Cohort. *Science* 287(5462):2398-2399. doi: 10.1126/ science.287.5462.2398. Available at http://science.sciencemag.org/content/287/5462/2398.

Gallagher, K., Magid, L., & Pruitt, K. (2016). The Educator's Guide to Student Data Privacy. *Ferpa|Sherpa*. Available at https://ferpasherpa.org/wp-content/uploads/2016/05/ EduGuide_DataPrivacy_516.pdf.

Gamoran, A., & Ferber, T. (2016, August 8). Why Would Have Thought: Bipartisan Poli-cymaking Through Evidence [Blog post]. Available at http://thehill.com/blogs/congress-blog/politics/290684-who-would-have-thought-bipartisan-policymaking-through-evidence?utm_content=bufferbfea7&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer.

Goel, V. (2014, June 29). Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry. *The New York Times*. Available at http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html?_r=0.

Goel, V. (2014, August 12). As Data Overflows Online, Researchers Grapple with Ethics. *The New York Times*. Available at http://www.nytimes.com/2014/08/13/technology/the-boon-of-online-data-puts-social-science-in-a-quandary.html?_r=1.

Guyette, C. (2015). How the EAA's Buzz Program exploited Detroit's Most Vulnerable Kids. *American Civil Liberties Union of Michigan*. Available at http://www.aclumich.org/article/guyette-how-eaas-buzz-program-exploited-detroits-most-vulnerable-kids.

Håberg, S. E., Trogstad, L., Gunnes, N., et al. (2013). Risk of Fetal Death After Pandemic Influenza Virus Infection or Vaccination. *New England Journal of Medicine* 368:333-340. doi: 10.1056/NEJMoa1207210.

Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The First Year of Open Online Courses. HarvardX and MITx Working Paper No. 1. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2381263.

IMS Global Learning Consortium. (n.d.). Caliper Analytics Background. Available at https://www.imsglobal.org/activity/caliperram.

Kaplan, S. (n.d.). Protecting Student Privacy in Research Projects. *Education New York*. Available at http://educationnewyork.com/files/040415-3.pdf.

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How Deep Is Knowledge Tracing? Ithaca, NY: Cornell University Library. Available at https://arxiv.org/abs/1604.02416.

Kirp, D.L. (2016). To Teach a Child to Read, First Give Him Glasses. *The New York Times*. Available at http://www.nytimes.com/2016/08/07/opinion/sunday/to-teach-a-child-to-read-first-give-him-glasses.html?_r=0&pagewanted=all.

Koedinger, K. R., Baker, R. S. J. d., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2011). A Data Repository for the EDM Community: The PSLC DataShop. In *Handbook of Educational Data Mining*, edited by C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.). Boca Raton, FL: CRC Press.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science* 36:757-798. doi: 10.1111/j.1551-6709.2012.01245.x.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional Complexity and the Science to Constrain It. *Science* 342(6161):935-937. doi: 10.1126/science.1238056.

Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013). Using Data-Driven Discovery of Better Student Models to Improve Student Learning. *Proceedings of Artificial Intelligence in Education*. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, edited by H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik. Pp. 421-430.

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data Mining and Education. *Wiley Interdisciplinary Reviews: Cognitive Science* 6:333-353. doi: 10.1002/wcs.1350.

Koedinger, K. R., McLaughlin, E. A., Jia, J. Z., & Bier, N. L. (2016). Is the Doer Effect a Causal Relationship?: How Can We Tell and Why It's Important. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (LAK '16). New York: ACM. 388-397. doi: http://dx.doi.org/10.1145/2883851.2883957.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Stamford, CT: META Group. Available at https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

LearnSphere. Available at http://learnsphere.org.

Leichty, R., & Leong, B. (2015). De-Identification and Student Data: Understanding De-Identification of Education Records and Related Requirements of FERPA. Washington, DC: Future of Privacy Forum.

Means, B., Padilla, C., & Gallagher, L. (2010). *Use of Education Data at the Local Level: From Accountability to Instructional Improvement*. Washington, DC: U.S. Department of Education. Available at http://eric.ed.gov/?id=ED511656.

Michelmore, K., & Dynarski, S. (2016). The Gap Within the Gap: Using Longitudinal Data to Understand Income Differences in Student Achievement. NBER Working Paper No. 22474. Available at http://www.nber.org/papers/w22474.

National Academy of Education. (2013). *Adaptive Educational Technologies: Tools for Learning, and for Learning About Learning,* edited by G. Natriello. Washington, DC: Author.

National Forum on Education Statistics. (2012). Forum Guide to Supporting Data Access for Researchers: A State Education Agency Perspective. NFES 2012-809. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

National Forum on Education Statistics. (2014). Forum Guide to Supporting Data Access for Researchers: A Local Education Agency Perspective. NFES 2014-801.Washington, DC: National Center for Education Statistics, U.S. Department of Education.

National Research Council. (1985). Sharing Research Data. Washington, DC: National Academy Press. doi: https://doi.org/10.17226/2033.

National Research Council. (1993). Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics. Washington, DC: National Academy Press. doi: https://doi.org/10.17226/2122.

National Research Council. (2000). Improving Access to and Confidentiality of Research Data: Report of a Workshop. Washington, DC: National Academy Press. doi: https://doi.org/10.17226/9958.

National Research Council. (2005). Expanding Access to Research Data: Reconciling Risks and Opportunities. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/11434.

Nicosia, M. (2016). "If I Was a Parent.. I would Opt Out": New NY Regents Chancellor Betty Rosa Comes Out Swinging. *The 74 Million*. Available at https://www.the74million.org/article/if-i-was-a-parent-i-would-opt-out-new-ny-regents-chancellor-betty-rosa-comes-out-swinging.

Nielsen, C. (2013). The Danish System for Access to Micro-data [PowerPoint slides]. Available at http://www.cbs.gov.il/twinning/c2/annex_c2.15_micro_data_and_confidentiality.pdf.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.

NORC at the University of Chicago. (n.d.). Data Enclave. Available at http://www.norc.org/Research/Capabilities/Pages/data-enclave.aspx.

NPRM (Notice of Proposed Rulemaking) for Revisions to the Common Rule: HHS Announces Proposal to Improve Rules Protecting Human Research Subjects. (2016). Available at http://www.hhs.gov/ohrp/regulations-and-policy/regulations/nprm-home.

Office for Human Research Protections (OHRP). (1993). *Institutional Review Board Guidebook*. Available at https://archive.hhs.gov/ohrp/irb/irb_preface.htm.

O'Rourke, E., Chen, Y., Ballweber, C., and Popovic, Z. (2016). Personalized Learning and Its Behavioral Impact on the Classroom Ecosystem (in submission).

Pardos, Z. A., & Heffernan, N. T. (2010). Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing. In *Proceedings of the 18th International Conference on User Modeling, Adaptation and Personalization*. Big Island, Hawaii. Pp. 255-266.

Pardos, Z. A., & Heffernan, N. T. (2011). KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model. In *Proceedings of the 19th International Conference on User Modeling, Adaptation and Personalization*. Girona, Spain.

Pardos, Z. A., Dailey, M. D., & Heffernan, N. T. (2011). Learning What Works in Its from Non-Traditional Randomized Controlled Trial Data. *International Journal of Artificial Intelligence in Education* 21:47-63. doi: 10.3233/JAI-2011-017.

Pardos, Z. A., Whyte, A., & Kao, K. (2016). moocRP: Enabling Open Learning Analytics with an Open Source Platform for Data Distribution, Analysis, and Visualization. *Technology, Knowledge and Learning* 21:75. doi: 10.1007/s10758-015-9268-2.

Perry, J. and Klopfer, E. (2014). UbiqBio: Adoptions and Outcomes of Mobile Biology Games in the Ecology of School. Computers in the Schools, Vol. 31, pp. 43-64.

Personal and Sensitive Data of 59,000 Charter School Students in California Leaked: Researcher. (2016, August 19). Available at https://www.databreaches.net/personal-and-sensitive-data-of-59000-charter-school-students-in-california-leaked-researcher.

Pitcan, M. (2016, June 13). Real Life Harms of Student Data: Moving Beyond the Hypothetical. *Points*. Available at https://points.datasociety.net/real-life-harms-of-student-data-956a30aaff32#.k0yu1gy2f.

P.L. 72-6217. Disclosure of Data; Requirements for Disclosure. (n.d.). Available at http://www.ksrevisor.org/statutes/chapters/ch72/072_062_0017.html.

Privacy Rights Clearinghouse. (2005, April 20). Data Breaches. Security Breaches 2005-Present. Available at http://www.privacyrights.org/data-breach.

PSLC Data Shop. Available at https://pslcdatashop.web.cmu.edu.

Rau, M. A., & Pardos, Z. A. (2012). Interleaved Practice with Multiple Representations: Analyses with Knowledge Tracing Based Techniques. In *Proceedings of the 5th annual International conference on Educational Data Mining*. Crete, Greece.

Reidenberg, J., Russell, N. C., Kovnot, K., Norton, T. B., & Cloutier, R. (2013). Privacy and Cloud Computing in Public Schools. New York: Fordham Center on Law and Information Policy.

S.1183. STEM Gateways Act. (2015). Available at https://www.congress.gov/bill/114th-congress/senate-bill/1183.

Schwartz, K. (2014). How Are Teachers and Students Using Khan Academy? *KQED News*. Available at http://ww2.kqed.org/mindshift/2014/05/06/how-are-teachers-and-students-using-khan-academy.

Sell, N., & Zolotova, R. (2016). Digital Toxic Waste (or Why Metadata Shouldn't Live Forever) [Blog post]. TechCrunch. Available at https://techcrunch.com/2016/07/13/digital-toxic-waste-or-why-metadata-shouldnt-live-forever.

Stanford Center for Education Policy Analysis. (n.d.). Stanford Education Data Archive. Available at https://cepa.stanford.edu/seda/maps.

Strauss, V. (2016). Reformers "Disrupted" Public Education. Now an Ivy League Dean Says the Consequences for Kids Can Be "Devastating." *The Washington Post*. Available at https://www.washingtonpost.com/news/answer-sheet/wp/2016/08/04/reformers-disrupted-public-education-now-an-ivy-league-dean-says-the-consequences-for-kids-are-devastating.

Surén, P., Bakken, I. J., Aase, H., Chin, R., Gunnes, N., Lie, K. K., Magnus, P., Reichborn-Kjennerud, T., Schjølberg, S., Øyen, A.S., & Stoltenberg, C. (2012, July). Autism Spectrum Disorder, ADHD, Epilepsy, and Cerebral Palsy in Norwegian Children. *Pediatrics* 130(1):e152-e158. doi: 10.1542/peds.2011-3217.

Surén, P., Roth, C., Bresnahan, M., et al. (2013). Association Between Maternal Use of Folic Acid Supplements and Risk of Autism Spectrum Disorders in Children. *JAMA* 309(6):570-577. doi: 10.1001/jama.2012.155925.

Tang, S., Peterson, J. C., & Pardos, Z. A. (2016). Modelling Student Behavior Using Granular Large Scale Action Data from a MOOC. Ithaca, NY: Cornell University Library. Available at https://arxiv.org/abs/1608.04789.

Tony. (n.d.). 6.002x Mongolia. Available at http://6002x.blogspot.com.

Ujifusa, A. (2014). State Lawmakers Ramp Up Attention to Data Privacy. *Education Week*. Available at http://www.edweek.org/ew/articles/2014/04/16/28data.h33.html?qs=data+privacy.

U.S. Department of Education. (2016). Family Education Rights and Privacy Act (FERPA). Available at http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html.

U.S. Department of Education Privacy Technical Assistance Center. Available at http://ptac.ed.gov.

U.S. Department of Education Privacy Technical Assistance Center. (2014). Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices. Available at https://tech.ed.gov/wp-content/uploads/2014/09/Student-Privacy-and-Online-Educational-Services-February-2014.pdf.

Vance, A. (2016). Trends in Student Data Privacy Bills in 2016. *Policy Update* 23(13):1-2. Alexandria, VA: National Association of State Boards of Education.

Vu, L., & Pardos, Z. A. (2016). Visualizing Online Learner Patterns. *ICML Visualization Workshop*. Available at https://icmlviz.github.io/assets/papers/21.pdf.

Wang, Y., Paquette, L., & Baker, R. (2014). A Longitudinal Study on Learner Career Advancement in MOOCs. *Journal of Learning Analytics* 1(3):203-206.

Wikipedia. (n.d.). Children's Online Privacy Protection Act. Available at https://en.wikipedia.org/wiki/Children%27s_Online_Privacy_Protection_Act.

Wikipedia. (n.d.). Family Educational Rights and Privacy Act. Available at https://en.wikipedia.org/wiki/Family_Educational_Rights_and_Privacy_Act.

Wikipedia. (n.d.). Natural Law. Available at https://en.wikipedia.org/wiki/Natural_Law.

Wikipedia. (n.d.). Protection of Pupil Rights Amendment. (n.d.).Available at https://en.wikipedia.org/wiki/Protection_of_Pupil_Rights_Amendment.

Wood, A., O'Brien, D., Altman, M., Karr, A., Gasser, U., Bar-Sinai, M., Nissim, K., Ullman, J., Vadhan, S., & Wojcik, M. (2014). Integrating Approaches to Privacy Across the Research Lifecycle: Long-Term Longitudinal Studies. *Berkman Center Research Publication* No. 2014-12. Available at http://ssrn.com/abstract=2469848 or http://dx.doi.org/10.2139/ssrn.2469848.

Yudelson, M. V. (2016). Individualizing Bayesian Knowledge Tracing: Are Skill Parameters More Important Than Student Parameters? *Proceedings of the 9th International Conference on Educational Data Mining*. Pp. 556-561.

Websites Pertaining to Big Data and Privacy

Student Data and Privacy Resources

- The Asilomar Convention for Learning Research in Higher Education
- Ask Before You App video
- Consortium of School Networking (CoSN)
    - Privacy Toolkit (free, but must create account to download)
    - Trusted Learning Environment Seal
- Common Sense Education
- Data Privacy MOOC
- Data Quality Campaign
- Educator's Guide to Student Privacy
- iKeepSafe
- Learning Through Teaching Resources
    - Berkman Center Digital Literacy Resource Platform
    - Berkman Center Safety, Privacy, and Digital Citizenship High School Curriculum
    - Fordham CLIP Volunteer Privacy Educators Program
- National Association of State Boards of Education (NASBE)
- Privacy Technical Assistance Center (PTAC, U.S. Department of Education)
- Student Data Principles: 10 Foundational Principles for Using and Safeguarding Students' Performance Information
- Student Data Privacy Communications Toolkit
- Student Data Privacy Consortium (Massachusetts Student Privacy Alliance Webpage)
- Wisconsin SEA Privacy Training

Sample Model Webpages from State and Local Educational Agencies

- Chesterfield County Public Schools Privacy Page
- Denver Public Schools Student Data Privacy Page
- Fairfax County Privacy Page
- Houston ISD Privacy Page
- Ventura County e-Safety Committee Task Force
- Wisconsin SEA Privacy Page