EVALUATION OF TEXAS HOME INSTRUCTION FOR PARENTS OF PRESCHOOL YOUNGSTERS PROGRAM ON READING AND MATH ACHIEVEMENT FOR GRADES K TO 8

Noor Amal S. Abdulaziz

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2019

APPROVED:

Robin K. Henson, Committee Chair and Chair of the Department of Educational Psychology
Darrell Hull, Committee Member
Smita Mehta, Committee Member
Cynthia Frosch, Committee Member
Randy Bomer, Dean of the College of Education
Victor Prybutok, Dean of the Toulouse Graduate School
Abdulaziz, Noor Amal S. *Evaluation of Texas Home Instruction for Parents of Preschool Youngsters Program on Reading and Math Achievement for Grades K to 8*. Doctor of Philosophy (Educational Psychology), August 2019, 173 pp., 9 tables, 12 figures, 4 appendices, references, 152 titles.

This study was intended to evaluate the impact of socioeconomically disadvantaged children’s participation in the Texas Home Instruction for Parents of Preschool Youngsters (TX HIPPY) Program on their school readiness and academic achievement. The study used a quasi-experimental design and applied full and optimal propensity score matching (PSM) to address the evaluation concern of the impact of the TX HIPPY program on HIPPY participants’ academic achievement compared to non-HIPPY participants. This evaluation targeted former HIPPY participants and tracked them in the Dallas ISD database through Grade Levels K-8. Data were obtained by administering Istation’s Indicators of Progress (ISIP) for kindergarten, TerraNova/SUPERA for Grades K-2, and State of Texas Assessments of Academic Readiness for math and reading (STAAR) for Grades 3-8. HIPPY and non-HIPPY groups were matched using propensity score analysis procedures. The evaluation findings show that the TX HIPPY program positively influences kindergarten students to start school ready to learn. The findings of math and reading achievements suggest that HIPPY children scored at the same level or higher than non-HIPPY children did on math and reading achievement, indicating that TX HIPPY program has achieved its goal of helping children maintain long-term academic success. However, the evaluation findings also indicated that the impact evaluation framework must be designed with attention to higher-level factors beyond academic achievement that influence children’s academic success.
Copyright 2019

by

Noor Amal S. Abdulaziz
TABLE OF CONTENTS

LIST OF TABLES .......................................................................................................................... v

LIST OF FIGURES ....................................................................................................................... vi

IMPACT EVALUATION OF TEXAS HOME INSTRUCTION FOR PARENTS OF
PRESCHOOL YOUNGSTERS PROGRAM ON READING AND MATH ACHIEVEMENT
FOR GRADES K TO 8 ................................................................................................................... 1

Introduction ......................................................................................................................... 1
   Educational Effects of Poverty .......................................................................................... 3
   Home Visitation Programs .............................................................................................. 4
   HIPPY Program .............................................................................................................. 6
   Program Evaluation of HIPPY ...................................................................................... 9

Purpose of the Present Study .............................................................................................. 12

Evaluation Question .......................................................................................................... 13

Methods ..................................................................................................................................... 13
   Participants ....................................................................................................................... 14
   Evaluation Design and Procedure ............................................................................... 15
   Measures ........................................................................................................................... 19
   Data Analysis ................................................................................................................... 22

Results ....................................................................................................................................... 28

Discussion .............................................................................................................................. 46
   TX HIPPY Program Evaluation ...................................................................................... 46
   Role of Propensity Score Analysis in the Evaluation ................................................... 64
   Evaluation Recommendations ....................................................................................... 68

Limitations and Future Research ......................................................................................... 71

Conclusion .............................................................................................................................. 73

APPENDIX A. EXTENDED LITERATURE REVIEW ................................................................. 74

APPENDIX B. R-CODE FOR PROPENSITY SCORES ANALYSIS ........................................ 97

APPENDIX C. R-CODE FOR INDEPENDENT SAMPLE t-TEST ON THE Z-SCORES ...... 155

APPENDIX D. IRB APPROVAL .............................................................................................. 159
LIST OF TABLES

Table 1. Detailed Comparison between the Previous and Current Study.......................... 13

Table 2. Detailed Demographic Characteristics for Former HIPPY Participants 2017-2018-Dallas
ISD (n = 1318) ......................................................................................................................... 16

Table 3. Missing Data Analysis Results .................................................................................... 23

Table 4. Variables Used in the Analysis .................................................................................... 29

Table 5. Covariate Balance Checking between the Two Groups on Each of the Balancing
Variables ....................................................................................................................................... 33

Table 6. Sample Size of HIPPY Participants and Non-HIPPY participants Before and After
Matching ....................................................................................................................................... 60

Table 7. Treatment Effect of TX HIPPY Program on Reading and Math Achievement for Each
Grade Level Resulted from the Fully Matched Sample............................................................ 63

Table 8. Treatment Effect of TX HIPPY Program on Reading and Math Achievement for Each
Grade Level Resulted from the Optimally Matched Sample .................................................... 66

Table 9. Descriptive Statistics and Independent Samples t-test Results Comparing HIPPY and
non-HIPPY participants on the Z-scores for each Standardized Assessment................................ 70
LIST OF FIGURES

Figure 1. HIPPY logic model........................................................................................................ 10
Figure 2. Evaluation framework. .................................................................................................. 18
Figure 3. Love plots before and after propensity score matching for Grade K. ....................... 50
Figure 4. Love plots before and after propensity score matching for Grade 1 ...................... 51
Figure 5. Love plots before and after propensity score matching for Grade 2 ...................... 52
Figure 6. Love plots before and after propensity score matching for Grades 3, 4, and 5........ 53
Figure 7. Love plots before and after propensity score matching for Grades 6, 7, and 8. ....... 54
Figure 8. Propensity scores distribution before and after the matching for Grade K. .......... 55
Figure 9. Propensity scores distribution before and after the matching for Grade 1 .......... 56
Figure 10. Propensity scores distribution before and after the matching for Grade 2 ........ 57
Figure 11. Propensity scores distribution before and after the matching for Grades 3, 4, and 5.. 58
Figure 12. Propensity scores distribution before and after the matching for Grades 6, 7, and 8.. 59
IMPACT EVALUATION OF TEXAS HOME INSTRUCTION FOR PARENTS OF PRESCHOOL YOUNGSTERS PROGRAM ON READING AND MATH ACHIEVEMENT FOR GRADES K TO 8

Introduction

Many interrelated family-level risk factors affect children's development in early childhood: low-income, unemployment, low education level, social isolation, single motherhood, and teenage parenthood (Brown, 2008, 2015; Brown & Lee, 2017; Johnson, Martinez-Cantu, Jacobson & Weir, 2012). Those families are at risk of fewer positive parent-child interactions during early childhood (Anthony, King, & Austin, 2011; Evans, 2004). Between 2009 and 2010, 25% of the children in the United States aged birth to 5 and more than 15 million children under age 18 years living in poverty, and about 6 million children were at risk of child maltreatment (Barton, 2016; DeNavas-Walt, Proctor, & Smith, 2010). Additionally, nearly half (47%) of the children living in the United States aged birth to 3 years suffer from challenging circumstances within their families, including low socioeconomic status and social complexities (Jiang, Ekono & Skinner, 2015).

Generally, growing up in poverty harmfully affects children’s future years, and the consequences become more severe as poverty increases (Korat & Haglili, 2007). Living in poverty during childhood poses risks to long-term developmental outcomes for socioeconomically disadvantaged children aged birth to 3 years (Anthony et al., 2011). Hence, there is a national urge to provide help for these families to support positive outcomes and greater parent involvement in children’s development and learning for limited education, low-income families (Barnett, 2011). Many studies have shown that participation in early childhood education programs has a positive impact on children because they experience benefits that

The Patient Protection and Affordable Care Act (ACA) provided a federal commitment to enhancing the living conditions of families living in poverty with young children by sponsoring evidence-based home visiting programs. The approval of a one and a half billion evidence-based home visitation initiative to the states over five years, known as the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) Program (Health Resources and Services Administration [HRSA], n.d.). The primary goal of MIECHV is to support families with young children and pregnant women in mastering skills and to obtain resources needed to promote development, health, and school readiness. Federal agencies recognized 17 home visiting intervention models in 14 states, including Healthy Families America (HFA), Home Instruction for Parents of Preschool Youngsters (HIPPY), Nurse-Family Partnership (NFP), Early Head Start (EHS), and Parents as Teachers (PAT). Home visitation programs are generally designed to enhance children’s health, and developmental outcomes together with parents’ capacities and skills (Barnett, 2011; Brown, 2008, 2015; Brown & Lee, 2017). Supported by federal policy, early childhood home visiting interventions have expanded across most of the United States in recent years (Lanier, Maguire-Jack, & Welch, 2015).

The nationally implemented models, including HIPPY, emphasize a commitment to the model’s fidelity and support rigorous evaluation of the programs’ effectiveness, which promotes better advocacy for strategies related to early childhood through home visiting (Barton, 2016; Boller, Strong, & Daro, 2010). Several comprehensive reviews and meta-analyses on the effectiveness of home visitation programs have shown success in improving children’s socioemotional development and parenting knowledge and skills (Carneiro, Meghir, & Parey,
However, early childhood education programs, including home visitation, could be unproductive if it failed to meet three criteria: value acceptance, technical feasibility, and cost-effectiveness (Kingdon, 2003). Hence, the magnitude to which early childhood intervention programs create enduring gains on children's socialization, cognitive development, and school success is a matter of debate when it comes to the feasibility and cost-effectiveness of the effect (Barton, 2016).

Educational Effects of Poverty

Children from low income, limited education, minority families, and single parent often enter school with social-emotional and health problems and limited language skills (Espinosa, 2007; National School Readiness, 2005; Manza, Hughes, Barnabas, Bracalielloa, & Ginsburg-Blockb, 2010). Low-income families are less likely to visit community libraries or have books at home, and they are at risk of higher stress and poor mental health (Korat & Haglili, 2007). As a result, their children have fewer words to use and fewer enriching language experiences than advantaged parents (Evans, 2004). Unfortunately, many low-income preschool children experience poor literacy skills and language due to resource unavailability and parents’ perspective on the importance of children's literacy development (Manza et al., 2010). The research also shows that children, at every level of their development, experience fewer problems and make stronger gains when their parents are knowledgeable about their school experiences (Brown, 2008, 2015; Brown & Lee, 2017).

The intervention programs for the socioeconomically disadvantaged children that aim at increasing children’s school readiness and averting later academic interruptions are more useful if they occur during preschool years (Brown, 2008, 2015). Also, many studies have shown that children who participate in early childhood education programs show progress that perseveres
during the course of their education and into early middle age (Barnett, 2011; Camilli, Vargas, Ryan, & Barnett, 2010) and early school readiness is linked to successful academic achievement (Barnett, 2011; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002). Moreover, the cognitive, socioemotional, and behavioral characteristics of school readiness skills in preschool and kindergarten aged children can be predictive of their academic achievement and attention to academic tasks in later grades (Duncan, Ziol-Guest, & Kalil, 2010; Reynolds, Temple, Ou, Arteaga, & Barry; 2011).

Home Visitation Programs

Early childhood education is critical for low-income children, as early learning influences later learning, and skills produce skills (Knudsen, Heckman, Cameron, & Shonkoff, 2006). Socioeconomically disadvantaged children are likely to continue in a similar pathway of poor development (e.g., delays in numeracy and reading) unless they come across new resources, opportunities, or interventions. Similarly, parents characteristics and experiences (e.g., education level, income, employment) that form home environments are challenging to change (Chase-Lansdale & Brooks-Gunn, 2014). The challenging family-related factors associated with family income, socioemotional and behavior skills are reduced when parents respond effectively to positive behavior, confidently express affection, and provide a high level of structure and instruction (Connell & Prinz, 2002; Gershoff, Aber, Raver, & Lennon, 2007; Yeung, Linver, & Brooks-Gunn, 2002). Furthermore, the child is less likely to do well when he/she returns from a motivating educational setting to a challenging family environment compared to a child who is exposed to enriching environments both in and outside the home (Chase-Lansdale & Brooks-Gunn, 2014).
Therefore, in order to effectively change the challenging environment of children who are socioeconomically disadvantaged, early childhood education programs should concurrently target the child and his/her home environment. Parent-child interaction and connection is the first framework for learning during early childhood; the interaction varies depending on the language proficiency, family income, and parents’ level of education (Bakermans et al., 2003). Although there are many delivery methods for early intervention programs, home visiting is the most convenient to the parents and the children given the fact that parents are the very first teachers of their children and home is the first classroom (Barton, 2016).

Many research studies have conducted to assess the long lasting effect of early childhood programs; Branett (1995) reviewed 36 studies of large-scale public programs to assess the long-lasting effect of early childhood programs on students from low-income families. The review primarily focused on the impact of early childhood intervention programs on children's cognitive development. Findings indicated that early childhood education programs could achieve short-term gains to children’s intelligence quotient (IQ) and long-term impacts on grade retention, school achievement, social adjustment, and placement in special education. High-quality family and environmental factors (e.g., low social risks, higher socioeconomic status, and positive home learning experiences) are likely to be associated with children's positive subsequent developmental and educational outcomes (e.g., literacy, social competence, and educational success) (Foster, Lambert, Abbott-Shim, McCarty, & Franze, 2005). Young children obtain an assortment of literacy development skills and capabilities, including phonological awareness, receptive vocabulary, story understanding, and print knowledge and concepts, which allow them to utilize advanced literacy skills (Dickinson & McCabe, 2001). The stability of children’s literacy skills in preschool can serve as a strong predictor of reading abilities to middle childhood
(Dickinson & McCabe, 2001). Early childhood education programs have positive short- and long-term effect on many child outcomes; including, higher levels of cognitive development and social skills, higher high school graduation rates, college attendance, higher income, and less involvement with the criminal justice system (Deming, 2009).

Despite the promising results of those model programs, some cognitive development studies indicate that there is no evidence that the sustainability of the benefits continues beyond two years (Barnett, 2011). The evaluations of larger-scale, state-level, and federal programs indicate a distinct pattern across studies concerning perseverance and cost-effectiveness of the effect (Barnett, 2011; Muschkin, Ladd, & Dodge, 2015; U.S. Department of Health and Human Services, 2010). Additionally, the impact evaluation studies using observational data share a challenging technical feature that their results are subject to selection bias (Steiner, Shadish, & Clark, 2010). Thus, the rigorous impact evaluation of early childhood programs is an urgent priority to maintain high-quality service and to ensure their sustainability.

HIPPY Program

This study is intended to evaluate the effectiveness of one of 17 federally recognized national home visitation models in 14 states. One of these is HIPPY. The core philosophy of HIPPY is to promote parent involvement in community life and school to boost the likelihood of fruitful school experiences (HIPPY USA, n.d.). The HIPPY mission is to empower parents of 3, 4, and 5-year-old children so they can be engaged in their children’s education as primary educators in the home to help their children be ready for success in school and to bridge the gap between early educational experiences and adult success (HIPPY USA, n.d.). The program targets communities recognized as socioeconomically disadvantaged parents of young children to help them gain more self-confidence to overcome their circumstances (HIPPY USA, n.d.).
The program provides resources for parents with the intent of creating healthier and sustainable families and increasing the protective factors against the impact of poverty.

Moreover, the program encourages community involvement by joining efforts from local non-profit organizations and school districts. The HIPPY program has been credited with increasing standardized test scores, and school readiness for HIPPY enrolled children compared to non-HIPPY children (Baker, Piotrkowski, & Brooks-Gunn, 1998; Hinkley, 2018; Palladino, 2016; Turner & Hinkley, 2017). However, the evaluation design in some of these studies is subject to selection bias. Enrollment in HIPPY is voluntary, and participation follows a structured process, where coordinators organize the home visitation, pre- and post-school readiness assessments, group meetings, and other program activities. Nationally normed, standardized assessments are conducted to assess the program’s effectiveness and generate statistical comparisons before and after the intervention.

The short-term outcomes of the HIPPY program are the expected changes in parents and children during the program’s life cycle (at least one year, although two years are preferred). This includes increasing the parents’ self-efficacy regarding engagement in their children’s education and relevant activities, improving the parent-child relationship, and improving the children’s pre-academic skills. The medium-term outcomes of the HIPPY program are the expected changes in the parents and children by the end of the program (after at least two years), including increasing the parents’ self-efficacy regarding involvement in their children’s education, setting high academic expectations for the children, becoming active in guiding the children’s educational experiences at home, improving family relationships, expanding children’s pre-academic environment, advocating for the children’s education, communicating with the children’s school, attending school events, and volunteering at school. These pre-
academic skills and knowledge in crucial learning domains enable children to be successful in kindergarten and start school ready to learn. The long-term outcomes of the HIPPY program include parents being expected to sustain involvement in the school learning system, increased family involvement in local communities, stronger parent-child relationships, and promotion of long-term success for children as a result of training and experiences received through the program.

Implementing the TX HIPPY program usually, encompass challenges related to many factors (e.g., child, family, cultural, staff, and resources) that can affect the way in which the program is implemented. These factors can promote program adaptation to meet stakeholders’ needs; however, too much adaptation may result in a loss of fidelity to the original logic model and theory of change. Assessing fidelity is essential in making sense of successful and unsuccessful outcome and verify whether the success or failure reflects the model design appropriateness and the implementation process. As stated by the HIPPY USA, the fidelity of implementation of the HIPPY program involves two parts (HIPPY USA, n.d.). The first includes self-assessment by the program’s sites carried on by a team including the site coordinator, a home visitor, a parent, and a representative from the sponsoring organization. Second, the program’s quality is monitored by a national HIPPY trainer through conducting file reviews, interviews, and observations of home visits. The HIPPY site earns accreditation status for three years when all the HIPPY model standards are met. If not, the site needs to develop a continuous improvement plan. Nevertheless, in the process of implementing this evaluation, it was noted that there is a need to guarantee that the program is being implemented as designed, learn about and disseminate information on effective and ineffective adaptations.
The TX HIPPY Logic Model (was developed by the evaluator of current study and approved by Texas HIPPY state office management) (Figure 1) presents a chain of causation of activities and outcomes that are expected to take place for parents and children who are HIPPY enrolled at ages 3, 4, and 5 years. The model has been developed as a forwarding “if…, then” method between the activities and outcomes that show a logically stated linear progression.

Program Evaluation of HIPPY

Intervention programs in the field of education are complex. They often include a diagnostic component to identify the needs along with a service component; this complexity increases the challenges for program evaluation due to the non-randomized assignment to the program (Bowden, Shand, Belfield, Wang, & Levin, 2017). The main goal of evaluating intervention programs in education is to identify the treatment effect (Mark & Henry, 2006).

The prime research concern in such an evaluation is whether the intervention improves educational outcomes. Program evaluation is basically an investigation of assumed cause and effect associations to answer the critical question of “To what extent can the difference observed in outcomes between the treatment and comparison group be attributed to the intervention, holding all other factors constant?” Causal inference in this setting implies the gain or loss detected in the outcome of the treated group that can be credited to the treatment variable (Rubin, 1986). The home visiting evaluation can be assessed through a systematic search process, screening, an assessment of program effectiveness, and a review of the evaluation quality (U.S. Department of Health and Human Services, 2010). This evaluation is meant to measure the impact of TX HIPPY on its long-term outcome of children’s academic achievement using a quasi-experimental design.
### Texas HIPPY Program – Logic Model

<table>
<thead>
<tr>
<th>Problem/Goal</th>
<th>Activities</th>
<th>Outputs</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home Visits: 30 week program includes weekly or biweekly home visiting:</td>
<td>Parent-Child Teaching Parent completes HIPPY activity packets and supplemental extension activities with the child according to instructions on a weekly basis.</td>
<td>Parent Learning: - Education and school terminology about child development and age appropriate expectations; - HIPPY’s guiding principles and teaching skills (parenting skills, child development, community resources, and school involvement); - How to use and care for educational materials; - How to initiate, monitor, and direct child’s pre-academic educational activities; - How to improve scheduling and time management skills.;</td>
<td>Complete at least years 4 &amp; 5 of the HIPPY curriculum; Parent Involvement increases; - Communication skills for discussing child’s educational activities; - Frequency of engaging in educational activities with child in home and community; - The use of HIPPY teaching skills; - Working with child on educational activities; - Connection with community and local school; - Comfort and interest in participating in school-related activities; - The quality of Parent-Child Interaction.</td>
</tr>
<tr>
<td>- Enroll the families in the program; - HIPPY model orientation; - Consent Form; - Conduct Pre-Assessments (Parent Involvement, BSRA-3, ASCQ-SE and PFS parent survey); - Deliver HIPPY packets for 30 weeks; give parent weekly activity, role-plays curriculum packets; and review the previous activity packet; - Gives parent other educational materials and resources; extension on activities, sets performance expectations for parent; - Conduct Post-assessments by the end of the interventions (Parent Involvement, BSRA-3, ASCQ-SE and PFS parent survey).</td>
<td>Social Interaction Parent has increased opportunities learning in a group setting and expanding their social circle.</td>
<td>Child School Readiness: - Acquires skills and values that display a predisposition to learning; pre-academic skills and knowledge in key domains; - Starts school ready to learn; - Home literacy environment improves.</td>
<td></td>
</tr>
<tr>
<td>Problem: Disadvantaged Parents (lack of education, poverty, social isolation, or other issues) are less likely to involve with their children’s education.</td>
<td>Group Meetings Program holds group meetings, during which:</td>
<td></td>
<td>Parent Self-efficacy: - Becomes active in guiding child’s educational experiences in the home; - Expands child’s pre-academic environment; - Parent involvement increases and family relationships improve; - Assumes an active role as child enters the formal academic environment by advocating for child’s education; - Communicating with child’s school; - Attending school events and volunteering in the school.</td>
</tr>
<tr>
<td>- help parents set goals for self and child; - Model behavior for parent; - Provide social-emotional support and encouragement to parent; - Build rapport with parent.</td>
<td>- Parents role play; provide parent with information about child learning and development, additional parenting information, educational materials, information, and resources. - Parents learn about school culture and organization.</td>
<td>Child skills: Child’s pre-academic skills improve.</td>
<td>Children’s Achievement: Long-term academic success. Family Involvement in local community activities is increased. Families have strong parent-child relationships.</td>
</tr>
</tbody>
</table>

**Figure 1.** HIPPY logic model.
In the last five years, the Dallas Independent School District (DISD) implemented evaluations of the Dallas HIPPY program to report the program outcomes (Hinkley, 2018; McEnturff, 2014; Palladino, 2016; Turner & Hinkley, 2017). The long-term outcome evaluation of the studies used quasi-experimental design and applied greedy (one to one) propensity score matching to create a counterfactual comparison group to control for selection bias. McEnturff (2014) tracked former HIPPY participants who were in Grades K to 8 for the school year 2013-2014 with a sample size of 1,483. The Iowa Test for Basic Skills (Logramos, Spanish version for reading) for math and reading for Grade K - 2, and State of Texas Assessments of Academic Readiness for math and reading (STAAR) for Grades 3 to 8 were used as the evaluation measures. The results indicated that former HIPPY participants outperform the non-HIPPY participants in all comparisons except for Grades K and 6 in reading and Grade 5 in math.

Palladino (2016), Turner and Hinkley (2017), and Hinkley (2018) used the same analysis and reporting methodology, with a sample size of 1,752 (Grades K-12), 1,266 (Grades K-5) and 1,231 (Grades K-3), respectively. Istation Indicators of Progress (ISIP) for kindergarten school readiness, TerraNova (Supera, Spanish version for reading) for math and reading for Grade K to 2, and State of Texas Assessments of Academic Readiness for math and reading (STAAR) for Grades 3 to 8 were used as the evaluation measures. The results indicated that former HIPPY participants outperform the non-HIPPY participants in all comparisons with non-statistically significant results except for K to 2 grades TerraNova and Supera assessments with small Cohen’s $d$ effect size between .21 and .26.

There is a possibility here for a great deal of improvement. Despite the infrequent use in the applied literature, optimal PSM method (optimal and full) have proven advantageous in reducing the observed bias and accurately estimating the treatment effect unlike the greedy
matching (one to one, one to k, and matching with a caliper) (Austin, 2014). Additionally, the methodology section of DISD studies indicated that either the analysis was conducted without pre-set balance checking criteria or checking the balance using the statistical significance testing, which is sensitive to sample size. Conducting a balance check is an important step to evaluate the effectiveness of the matching; this is usually done by comparing the standardized mean difference and variance ratio between the groups on the covariates for pre and post-matching. Furthermore, there is no indication in the DISD studies on the sensitivity test. Conducting sensitivity analysis in observational studies is a crucial step to checking for hidden bias; it tells whether the treatment effect estimation is accurate with a low standard error or if the estimation is confounded with other factors that might be overlooked in the statistical model. Table 1 provides a detailed comparison between the previous and current studies.

Purpose of the Present Study

HIPPY is a school preparation, parent involvement, program that supports socioeconomically disadvantaged families prepare their children for success in school. The enrolled participants are the parents of 3 to 5-year-old children from communities recognized as having limited education, low-income, or lack social support to help them gain more self-confidence to overcome their circumstances. This external evaluation was intended to measure the impact of TX HIPPY on its long-term outcome of academic success at the Texas Dallas site on children’s academic achievement for the school year 2017-2018 using a quasi-experimental design. The findings of the evaluation demonstrated TX HIPPY accountability, which highlights the program’s contribution to achieving long-term academic success for HIPPY enrolled children.
Table 1

**Detailed Comparison between the Previous and Current Study**

<table>
<thead>
<tr>
<th></th>
<th>(McEnturff, 2014)</th>
<th>(Hinkley, 2018; Palladino, 2016; Turner &amp; Hinkley, 2017)</th>
<th>Current evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Long-term reading and math outcomes for former HIPPY participants.</td>
<td>Long-term reading and math outcomes for former HIPPY participants.</td>
<td>Long-term reading and math outcomes for former HIPPY participants.</td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td>Quasi-experimental</td>
<td>Quasi-experimental</td>
<td>Quasi-experimental</td>
</tr>
<tr>
<td><strong>Matching</strong></td>
<td>Greedy PSM (1:1)</td>
<td>Greedy PSM (1:1)</td>
<td>Optimal PSM (Optimal and Full)</td>
</tr>
<tr>
<td><strong>Balance checking</strong></td>
<td>Univariate and multivariate Compares pre and post without setting criteria.</td>
<td>Univariate and multivariate Compares pre and post using the statistical significance testing, which is sensitive to sample size.</td>
<td>Compares pre and post propensity score matching using the standardized mean difference and variance ration.</td>
</tr>
<tr>
<td><strong>Measure</strong></td>
<td>ITBS, Logramos, and STAAR</td>
<td>Istation, TerraNova(Supera), and STAAR</td>
<td>Istation, TerraNova(Supera), and STAAR</td>
</tr>
<tr>
<td><strong>Treatment Effect</strong></td>
<td>( t )-test and effect size to report ATT</td>
<td>( t )-test and effect size to report ATT</td>
<td>Optimal: ( t )-test and effect size to report ATT. Full: regression adjustment within each stratum and then aggregate the treatment variable across all strata to report ATT.</td>
</tr>
<tr>
<td><strong>Sensitivity Analysis</strong></td>
<td>NA</td>
<td>NA</td>
<td>Rosenbeam framework 1991, 2015 to detect the hidden bias.</td>
</tr>
</tbody>
</table>

**Evaluation Question**

The evaluation employed a matched comparison group of non-HIPPY participants to provide an answer to; what is the impact of the TX HIPPY program on HIPPY participants’ academic achievement compared to non-HIPPY participants for the school year 2017 - 2018?

**Methods**

The study participants were located in Dallas, Texas. In 1988, the TX HIPPY program
was first piloted at this site by serving 14 families in West Dallas. Dallas HIPPY is currently serving over 1,000 families. The program provides intervention for 3, 4, and 5 years old children, who are from low-income and limited education families.

Participants

The evaluation study took place within the DISD and included all program and non-program participants in the district in Grades K to 8. Using optimal matching propensity score algorithms, the study selected former HIPPY enrolled children who were in Grades K to 8 for the school year 2017-2018 as the treatment group, along with a matched group of non-HIPPY students from the same grade levels as the comparison group. Table 2 shows the detailed demographic characteristics for former HIPPY participants in the DISD.

The treatment group consisted of Dallas ISD students in Grades K to 8 who participated in the TX HIPPY program when they were pre-school age ($n = 1,318$). Participants served were 646 (49%) female and 672 (51%) male. The primary language of the majority was Spanish (1,062; 78%), with most others speaking English (205; 16%). The majority were born in the United States (96%), and the rest from other countries (4%). Most were eligible for Limited English Proficiency program (79%) with the rest as unreported (12%), exited (2%), or not eligible (7%). The majority of the participants were Hispanic (1,195; 91%), and most others were African American (104; 8%). The majority of the participated children were considered to be 'at risk' due to factors such as low income and low educational achievement (1,085; 82%). There were 89 (7%) students eligible for the special education program, and, 241 (18%) enrolled in the talented and gifted programs. Most of the participants (550; 42%) were eligible for free and reduced lunch. Seventeen percent (17%) have a disability, 1% enrolled in Dyslexia Reading program, and 1% were identified as homeless.
The comparison group participants were also Dallas ISD students from the same grade levels as the program group participants but did not receive the program \((n = 61,211)\). The majority of non-HIPPY participants were Hispanic (73%), with the reminder African American (21%), White (4%), or other (2%). The sample was 49% female and 51%, male. Forty-six percent (46%) reported that their primary language is Spanish, 46% English, and 7% as English/Spanish. The majority were born in the United States (96%), and the rest from other countries (3%) and Honduras (1%). Most were eligible for Limited English Proficiency program (51%) with the rest as unreported (40%), exited (2%), or not eligible (7%). Eight percent (8%) were eligible for special education. Fourteen percent (14%) were enrolled in the Talented and Gifted program. The majority were from low socio-economic families (85%) with 66% identified as at-risk of academic failure due to different factors. Twelve percent (12%) have a disability, 3% enrolled in Dyslexia Reading program, and 1% were identified as homeless.

Evaluation Design and Procedure

HIPPY participants are recruited from communities recognized as having limited education, low-income, or lacking a network of social support. TX HIPPY activities involve home visits, which function as the main scheme within the program. The program delivers the intervention for 30 weeks during the program year. The focus of home visiting activities is role-playing and modeling teaching behavior to introduce parents to methods and materials that will prepare their children for school and encourage them to become actively involved in their children’s learning process, with consideration for the parents’ native language and literacy levels (Westheimer, 2003). Following each home visit, the parents are required to spend some time (15-20 minutes) on the activity packets with their children in order to achieve the program’s goal of providing developmentally appropriate pre-academic instruction.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Pre-K</th>
<th>K</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Former HIPPY Participants</td>
<td>3%</td>
<td>23%</td>
<td>12%</td>
<td>15%</td>
<td>13%</td>
<td>6%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>100%</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>8%</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>0%</td>
<td>22%</td>
<td>11%</td>
<td>13%</td>
<td>12%</td>
<td>6%</td>
<td>6%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>91%</td>
</tr>
<tr>
<td>Other</td>
<td>-</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>-</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>3%</td>
<td>22%</td>
<td>11%</td>
<td>13%</td>
<td>12%</td>
<td>6%</td>
<td>6%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>91%</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>0%</td>
<td>2%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>1%</td>
<td>9%</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1%</td>
<td>11%</td>
<td>6%</td>
<td>8%</td>
<td>6%</td>
<td>3%</td>
<td>3%</td>
<td>4%</td>
<td>4%</td>
<td>2%</td>
<td>49%</td>
</tr>
<tr>
<td>Male</td>
<td>1%</td>
<td>12%</td>
<td>6%</td>
<td>8%</td>
<td>7%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>4%</td>
<td>3%</td>
<td>51%</td>
</tr>
<tr>
<td>Primary Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>0%</td>
<td>3%</td>
<td>2%</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>16%</td>
</tr>
<tr>
<td>English/Spanish</td>
<td>0%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>6%</td>
</tr>
<tr>
<td>Spanish</td>
<td>2%</td>
<td>18%</td>
<td>9%</td>
<td>11%</td>
<td>10%</td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
<td>4%</td>
<td>7%</td>
<td>78%</td>
</tr>
<tr>
<td>Country of Birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>-</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>United States</td>
<td>2%</td>
<td>23%</td>
<td>11%</td>
<td>15%</td>
<td>13%</td>
<td>6%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>97%</td>
</tr>
<tr>
<td>Special Education Service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>1%</td>
<td>7%</td>
</tr>
<tr>
<td>No</td>
<td>2%</td>
<td>22%</td>
<td>11%</td>
<td>14%</td>
<td>12%</td>
<td>6%</td>
<td>6%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>93%</td>
</tr>
<tr>
<td>Gifted &amp; Talented Participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>18%</td>
</tr>
<tr>
<td>No</td>
<td>3%</td>
<td>23%</td>
<td>10%</td>
<td>12%</td>
<td>10%</td>
<td>5%</td>
<td>4%</td>
<td>4%</td>
<td>6%</td>
<td>4%</td>
<td>82%</td>
</tr>
<tr>
<td>Low SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2%</td>
<td>19%</td>
<td>11%</td>
<td>13%</td>
<td>13%</td>
<td>6%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>91%</td>
</tr>
<tr>
<td>No</td>
<td>0%</td>
<td>4%</td>
<td>1%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>-</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>9%</td>
</tr>
<tr>
<td>Dyslexia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>No</td>
<td>3%</td>
<td>23%</td>
<td>12%</td>
<td>15%</td>
<td>13%</td>
<td>6%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>99%</td>
</tr>
<tr>
<td>At-Risk Children</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3%</td>
<td>19%</td>
<td>10%</td>
<td>12%</td>
<td>10%</td>
<td>5%</td>
<td>6%</td>
<td>6%</td>
<td>7%</td>
<td>4%</td>
<td>82%</td>
</tr>
<tr>
<td>No</td>
<td>0%</td>
<td>4%</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>18%</td>
</tr>
<tr>
<td>Homelessness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>No</td>
<td>3%</td>
<td>23%</td>
<td>12%</td>
<td>15%</td>
<td>13%</td>
<td>6%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>99%</td>
</tr>
<tr>
<td>Free/Reduced Lunch Eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible for free/reduced lunch</td>
<td>-</td>
<td>-</td>
<td>0%</td>
<td>11%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>7%</td>
<td>5%</td>
<td>42%</td>
<td></td>
</tr>
<tr>
<td>Not Identified as Eligible</td>
<td>-</td>
<td>-</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>-</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>Other Disadvantaged Situations</td>
<td>2%</td>
<td>19%</td>
<td>11%</td>
<td>13%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>49%</td>
</tr>
<tr>
<td>Not available</td>
<td>0%</td>
<td>4%</td>
<td>1%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>8%</td>
</tr>
</tbody>
</table>
Furthermore, group meetings with the parents are held a minimum of six times a year to increase parents’ social interaction and to encourage them to learn within the group. Nationally normed, standardized assessments are conducted to evaluate the program effectiveness and generate comparisons of results for before and after the intervention. Home visitors’ training is another part of the program; home visitors are required to be educated under similar guidelines on the use of the curriculum and assessments to ensure that HIPPY enrolled children administer the same curriculum in the same order. Additionally, a system of model fidelity is applied to ensure that program sites are implementing the program model as designed.

The HIPPY participants are not randomly selected from the population of eligible families. The differences in measured outcome indicators between HIPPY participants and non-participants may not be credited to the program alone. This evaluation was intended to estimate the TX HIPPY program effectiveness on math and reading achievement for former participants. Primary data was collected through standardized tests administered by Dallas ISD for the school year 2017 - 2018. Dallas ISD provided the source of the participants' demographic data and the standardized test scores for former HIPPY participants and non-HIPPY participants via a de-identified list of HIPPY and non-HIPPY students and their data through a data-sharing agreement.

Former HIPPY participants were tracked (to allow the analysis of long-term outcomes) and matched with non-HIPPY participants as a comparison group for the purpose of creating a counterfactual comparison group using the PSM method to evaluate the program’s impact. The use of the right selection mechanism is the key to produce a balance between the HIPPY enrolled participants and non-HIPPY participants’ groups.

The most popular algorithm in the propensity score matching is the one-to-one greedy
matching, which controls one participant from the comparison group to match with one treated participant based on the propensity score value of baseline covariates. In the sense that the matching does not, for example, pair one treated Hispanic male with one non-treated Hispanic male; it matches based on one number (propensity score), which makes it associated with a significant sample selection bias. The paired participants are considered as if they were randomly assigned to the groups, which is not the case due to the selection bias. Hence, making comparisons between them to estimate the treatment effect is unfair. Therefore, optimal PSM based on some characteristics was to be employed to select the sample frame to prevent selection bias and make a reasonable causal attribution about the TX HIPPY program impact (Caliendo & Kopeinig, 2008; Rosenbaum, 2002; Rubin, 2006; Wang et al., 2013). The counterfactual comparison group helped provide a projection of what might have happened to HIPPY enrolled children had they not received the program. Hence, the comparison between program participants and non-participants that share the same characteristics will lead to an estimate of the TX HIPPY effect on the participants. Figure 2 shows the evaluation framework.

*Figure 2. Evaluation framework.*
As the evaluation used an existing database, the approved IRB application for conducting data collection and analysis for the program was sufficient for conducting the study. However, the process of accessing the Dallas ISD database required a research proposal application to the Dallas ISD Research Review Board. The confidentiality of the data was maintained; only the research team had access to the data. The study used de-identified information for individual subjects. As the evaluation used an existing database, there was no direct contact with the study participants.

Measures

Standardized assessments have been increasingly required by legislation to track individuals’ success in school (Brigman et al., 2015). Texas is required, as are all other states, to conduct annual assessments to monitor academic progress in standards defined by the state. The outcome measures of the TX HIPPY program are as follows:

*Istation Indicators of Progress (ISIP)*

This computer adaptive testing system is designed to continuously monitor the progress of reading growth for Grades Pre-K to 3 students. ISIP Early Reading and Early Math assessments for Grades Pre-K to 2 measures skills considered as most productive of reading and math success. The assessment serves as a diagnostic tool for the children’s school readiness in Grade K. ISIP early reading measures the child’s placement in the critical early reading skills; alphabetic knowledge and skills, comprehension, fluency, phonemic awareness, and vocabulary. ISIP early math assessment measures the children’s mathematical abilities in foundational skills; number sense, mathematical reasoning, geometry, operations, algebra, algebraic thinking, measurement, data analysis, probability and statistics, and personal financial literacy. The study used the ISIP assessment score to compare HIPPY participants and non-HIPPY participants for
kindergarten children on the school readiness measure. The assessment is intended to categorize children vulnerable to reading difficulties, automatically gives progress monitoring of skills that are indicators of later reading success and gives instant connection of assessment data to children’s learning needs that endorses differentiated instruction (Mathes, Torgesen, & Herron, 2016). Data from ISIP early reading and math are valid and reliable for students in Grades Pre-K to 3; the internal consistency coefficient was .891, suggesting that ISIP early reading scores were reliable (Mathes et al., 2016). The scores also demonstrated moderate to strong convergent and discriminant validity (Mathes et al., 2016).

*TerraNova and Supera*

This assessment is a set of norm-referenced achievement measures designed to provide achievement scores on students’ academic progress in reading, language arts, and math for Grades K to 2. These assessments use selected-response items to offer comprehensive comparative and diagnostic information on students’ academic progress. TerraNova and Supera measure the children’s abilities in foundational reading skills; analyze text, basic understanding, evaluate and extend meaning, reading and writing strategies, oral comprehension, and introduction to print. TerraNova math assessment is used to measure important mathematical skills; computation and numerical estimation, number and number relations, measurement, operation concepts, geometry, data analysis, statistics, and probability, patterns, functions, algebra, problem solving and reasoning, and communication. The study used the assessment score to compare HIPPY participants and non-HIPPY participants for Grades K to 2 on reading and math achievement. TerraNova and Supera have long-lasting and comprehensive data on validity and reliability (Data Recognition Corporation, 2016). The high internal consistency
coefficient of .80 speaks to appropriate reliability. The data also demonstrated strong convergent and discriminant validity (Data Recognition Corporation, 2016).

**State of Texas Assessments of Academic Readiness (STAAR)**

This state-mandated assessment consists of a sequence of standardized tests applied in public schools to assess students’ achievements, and knowledge, writing, and math. The assessment measures the fundamental reading skills; vocabulary acquisition and use, word relationships, structural analysis, context clues, real-life word connections and applications, vocabulary in context, antonyms, multiple-meaning words, synonyms, word reference materials, and figures of speech. Math assessment is used to measure critical mathematical skills; numerical estimation, number and number relations, operation concepts, geometry, data analysis, statistics, and probability, patterns, functions, algebra, and measurement. The study used the assessment scores to compare HIPPY participants and non-HIPPY participants in Grades 3 to 8 on math and reading achievement. Scores on this broadly used assessment have been previously validated (Texas Education Agency, 2016). Internal consistency reliability estimates range between .89 and .93 for math and reading Grades 3 through 8. The data also demonstrated strong convergent and discriminant validity (Texas Education Agency, 2016).

Standards-based education refers to systems of teaching, assessment, grading, and reporting that are hinge on students representing mastery or understanding of the skills and knowledge they are anticipated to learn as they progress through their education. As others have articulated (e.g., Bancroft, 2010; Gere et al., 2014), standardized assessments have limitations; they focus on measuring cognitive aspects overlooking other essential qualities necessary for children’s success. In addition, the majority of TX HIPPY children are English learners with different levels of language proficiency and their academic achievement performance was
compared relative to the English version test-takers in the normative sample; while they must be compared to a Spanish-speaking normative sample. It is possible that not using a Spanish normative sample could influence the results.

The following data were collected from the Dallas ISD database to conduct the study: (a) Istation Indicators of Progress (ISIP) for kindergarten children, Fall 2017 data was used to compare the kindergarten readiness of HIPPY versus non-HIPPY groups; (b) TerraNova(Supera) for Grades K to 2, Spring 2018 data was used to compare spring achievement results of HIPPY versus non-HIPPY groups for reading and math (Supera is only for reading); (c) STAAR for Grades 3-8, Spring 2018 data was used to compare achievement scores of HIPPY versus non-HIPPY groups on reading and math.

Data Analysis

The Dallas ISD database consisted of ISIP scores to measure the school readiness for kindergarten children, TerraNova(Supera) and STAAR scores for each of the subject area tests (reading and math) for Grades K to 8. Data cleaning was conducted for each assessment; the amount of missing data in the sample varied across each variable from .11% for TerraNova math scores to 83% for disability status. The data entry of the disability status covariate was coded to identify the children with disability and leave the other children code as blank. Table 3 presents the missing data analysis results. The missing data issues were addressed by creating 100 multiple imputed datasets for the analysis. Contemporary research on missing data imputation suggests that (as a rule of thumb) analyzing 20 datasets would maximize the analysis power; however, using more imputations is even better (Enders, 2010). The Statistical Package for Social Sciences (SPSS) software was used to conduct multiple imputation procedures. The R programming syntax (Appendix 1) was then implemented for the imputed data. For each grade
and subject matter, the 100 sets of parameter estimates, standard errors, and p values in each level of the analysis were pooled and reported as one estimate.

**Propensity Score Matching**

The propensity score matching (PSM) approach has become widely applied in the field of evaluation to estimate treatment effects (Caliendo & Kopeinig, 2008). The propensity score is defined as a participant’s probability of receiving a treatment based on a set of observed covariates; those covariates are estimated at baseline to control for selection bias. The propensity score is a method of establishing a similar distribution of confounders between treatment and comparison groups, thereby increasing the comparability between the groups (Wang et al., 2013). Matching participants that have close to identical propensity scores to equate treatment and comparison groups is the most common method (Rosenbaum, 2002; Rubin, 2006), comparing outcomes between treated and untreated participants to which treatment selection is not confounded with measured or unmeasured baseline covariates permits the study to attain an unbiased estimate of the treatment effect. Any difference in the outcome of interest can be credited to the effect of treatment when there is no systematic difference in measured or unmeasured baseline covariates between treated and untreated participants (Holmes, 2014).

**Table 3**

**Missing Data Analysis Results**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Missing N</th>
<th>Percent</th>
<th>Valid N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability Status</td>
<td>53438</td>
<td>83.43%</td>
<td>10627</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Score - TerraNova</td>
<td>9947</td>
<td>51.3%</td>
<td>9457</td>
<td>1135.46</td>
<td>97.138</td>
</tr>
<tr>
<td>Reading Score - STAAR</td>
<td>39786</td>
<td>44.91%</td>
<td>48731</td>
<td>1484.75</td>
<td>294.93</td>
</tr>
<tr>
<td>Math Score - STAAR</td>
<td>38278</td>
<td>43.22%</td>
<td>50239</td>
<td>1589.23</td>
<td>617.09</td>
</tr>
<tr>
<td>English Language Proficiency</td>
<td>24970</td>
<td>39.02%</td>
<td>39095</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>Missing N</td>
<td>Percent</td>
<td>Valid N</td>
<td>Mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>-------</td>
<td>---------------</td>
</tr>
<tr>
<td>Free or Reduced Lunch Status</td>
<td>8156</td>
<td>12.65%</td>
<td>55909</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country of Birth</td>
<td>127</td>
<td>.21%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Score - TerraNova</td>
<td>23</td>
<td>.11%</td>
<td>19381</td>
<td>787.09</td>
<td>288.381</td>
</tr>
</tbody>
</table>

Note. (a) Maximum number of variables shown: 25 (b) Minimum percentage of missing values for the variable to be included: .1%.

The propensity score analysis yields an unbiased treatment effect if the assumption of strong ignorable treatment effect holds (Rosenbaum & Rubin, 1983). Strong ignorability is defined as the lack of confoundedness between the experiment and the control groups, which indicates that no factors are confounding the relationship between the treatment and the outcome. This assumption is important in achieving a high-quality pretest measure that can be evaluated by conducting the sensitivity analysis (Shadish, Clark, & Steiner, 2008; Steiner et al., 2010). PSM is proven to be a reliable method in observational and quasi-experiment studies, but it has some limitations (Shadish & Steiner, 2010). First, this method might affect the study’s sample size due to the selection specifications. Second, the estimation bias can be extensive if the evaluator fails to control for important covariates in data from observational and quasi-experiment studies (Shadish & Steiner, 2010). This evaluation is intended to use the optimal full matching technique, which is proven effective in reducing the bias and achieving internal (balancing quality checking) and external validity (through the usage of the whole sample frame) (Green & Stuart, 2014). Its effectiveness derived from the fact that it is a combination of matching, weighting, and stratification techniques. This matching algorithm divides the sample frame into strata based on the propensity scores for which participants are only comparable within each stratum. The advantage of the stratification is that strata-level characteristics are automatically balanced, but the participants-level characteristics may remain imbalanced (Leite, Aydin, & Gurel, 2018).
**Propensity Score Estimation**

The propensity score is a method of establishing a similar distribution of the covariates between treatment and comparison groups, thereby increasing the comparability between the groups (Wang et al., 2013). The convenience sample selection process started for non-HIPPY participants using optimal full and optimal PSM based on specific demographic characteristics grounded in the literature that are likely to influence the treatment selection. The set of the child demographic characteristics included race, sex, English proficiency, child’s primary language, bilingual program participation, country of birth, special education status, age in days, gifted program status, free or reduced lunch status, low socioeconomic status, disability status, dyslexia, at-risk child, and homelessness. Consequently, a logistic regression model was used on the multiply imputed data sets to estimate the probabilities (propensity scores) of group membership for all participants (Enders, 2010; Guo & Fraser, 2010). The resulting coefficients, standard errors, \( p \) values, and \( F \) values were then pooled by calculating the weighted average.

**Propensity Score Application**

The estimated scores were then used to make the treatment and comparison groups similar through applying the matching technique (Guo & Fraser, 2010; Holmes, 2014). A conditioning strategy was deployed to produce groups based on the mean similarity and distribution of propensity scores. This was done using optimal and optimal full propensity score matching using the R statistical program with the `optmatch` package (Hansen & Klopfer, 2006), `MatchIt` package (Ho, Imai, King, & Stuart, 2011), together with the `cobalt` package for multiply imputed data (Greifer, 2019). Optimal full matching employs of all participants in the sample by establishing a series of non-overlapping matched subclasses in which each set has either one treated participant and multiple comparison participants or one comparison participant and
multiple treated participants (Stuart, 2010; Stuart & Green, 2008). This method has proven to be successful at reducing the bias of the observed confounding variables by 99% (Stuart, 2010). Optimal full matching forms pairs of treated and comparison participants in a way that minimizes the average within-pair difference in the propensity score (Austin & Stuart, 2015).

**Propensity Score Matching Evaluation**

The newly matched sample was evaluated to examine the effectiveness of conditioning and to measure the magnitude of the bias and any improvement after PSM. Internal validity was evaluated to see whether the covariates were balanced, the standardized mean difference threshold of .1 and variance less than two on each covariate indicated that the groups were matched effectively (Rubin, 2001). The standardized mean difference between .1 and .25 indicate that the propensity score model needed to be re-specified by excluding the imbalanced covariates (Guo & Fraser, 2010; Holmes, 2014). External validity was examined as well by checking the similarity between the propensity scores distributions of the matched HIPPY and non-HIPPY groups.

**Treatment Effect for the Optimally Matched Samples**

The treatment effect of the optimally matched sample followed the proposed outcomes framework by Rubin (1974), where each participant has a pair of possible outcomes: \( Y_i(0) \) and \( Y_i(1) \), the outcomes of untreated and treated, respectively. Each participant can be either treated or untreated. Let \( W \) be a variable referring to the treatment received (\( W = 0 \) for non-treatment individuals vs. \( W = 1 \) for treatment individuals). Hence, only one outcome (\( Y \)), is observed for each participant. \( Y_i \) is defined to be: \( Y_i(0) \) if \( W_i = 0 \) and to be \( Y_i(1) \) if \( W_i = 1 \) (alternatively, \( Y = WY(1) + (1 - W)Y(0) \)). Then, the average treatment effect for the treated (ATT), expressed as \( E[Y(1) - Y(0)|W = 1] \), is the average effect of treatment on
those participants who eventually received the treatment (Imbens, 2004). Accordingly, the mean difference ($t$-test) and effect size (Cohen’s d) was conducted to estimate the treatment effect.

*Treatment Effect for the Fully Matched Samples*

The analysis used a different statistical model; the analysis started with adjusting the imbalance within each stratum by regressing the outcome measure (test scores) on the treatment variable. The coefficient of the treatment variable was then aggregated across all the strata and weighted by the total number of the treated to estimate the average treatment effect on the treated (Stuart, 2010). The analysis was conducted using *R programming* software with the *Zelig* package (Imai, King, & Lau, 2011).

*Treatment Effect for Each Assessment across the Grades*

The attrition rate is high for Grades 4 to 8 with sample size ranges from 85 to 109. The power of the statistical test was assessed using the effect sizes resulted from the previously mentioned studies conducted by DISD; the analysis showed that at least 253 participants are needed for each grade level to prove the effect size of .25. In order to overcome this limitation, the study converted the standard scores for each assessment across the grades to z-scores to have them all on the same scale and conducted independent sample $t$-test and effect size to assess the effect of TX HIPPY program (Hull, Hinerman, Ferguson, Chen, & Näslund-Hadley, 2018). The treatment effect estimation methodologies were conducted separately for each grade level and each outcome measure (ISIP, TerraNova(Supera), and STAAR) as the statistical model for the outcome analysis.

*Sensitivity Analysis to Detect Selection Bias*

The last step in the analysis was to test the sensitivity of the findings concerning the
deviations from the pre-identified assumption and to address the limitation of an inability to adjust for unobserved covariates (hidden covariate bias). The analysis was conducted using the R statistical program with the `rbounds` package (Rosenbaum, 2001), started with selecting a sequence of odd ratio (Γ) values beginning at 1, then conducted the Wilcoxon signed rank test (for continuous outcomes) for each ratio (Γ). The upper bound of the p-value of the test statistic is examined for change over the sequences of Γ values. The treatment effect is considered sensitive when the Γ value changed from statistically significant to non-significant, and very robust to hidden bias if no change has occurred.

**Results**

The analysis was repeated for each grade level. The propensity scores were estimated by employing logistic regression of the HIPPY participation variable on the 15 baseline covariates (defined in Table 4) of the participants’ likelihood of receiving the TX HIPPY program. The evaluation created a matched sample by matching HIPPY participants and non-HIPPY participants. The optimal matching algorithms with replacement were employed to form pairs of HIPPY and non-HIPPY participants. The optimal full and optimal matching on the propensity scores have eliminated many of the observed differences in means and variance between HIPPY and non-HIPPY participants. Table 5 shows the covariate balance checking between the two groups on each of the matching variables for math and reading Grades K to 8, and school readiness for Grade K. Balance checking between the two groups was conducted to test whether or not this process generated a comparable comparison group. The matching on the propensity scores created a matched sample in which the distributions of the HIPPY and non-HIPPY participants are very similar. Additionally, the variances ratios of the continuous covariates were closer to one after matching, indicating that the matching was effective in eliminating the
observed bias. For instance, the most substantial standardized mean difference at baseline for kindergarten participants was for the primary language $d = .88$; this was reduced to $d = .02$ after matching, and the variance ratio of the unmatched sample between HIPPY and non-HIPPY for the race covariate was 7.51 and was decreased to 1.34 after matching.

Table 4

*Variables Used in the Analysis*

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIPPY</td>
<td>Numeric-Categorical</td>
<td>Yes = 1, No = 0</td>
</tr>
<tr>
<td>Race</td>
<td>Numeric - Nominal</td>
<td>American Indian or Alaskan = 0, Asian = 1, African American = 2, Hispanic / Latino = 3, Native Hawaiian or other pacific island = 4, Not available = 5 two or more race = 6, White = 7</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Numeric - Nominal</td>
<td>Not Hispanic or Latino = 0, Hispanic or Latino = 1</td>
</tr>
<tr>
<td>Sex</td>
<td>Numeric - Nominal</td>
<td>Female = 0, Male = 1</td>
</tr>
<tr>
<td>Limited English Proficiency</td>
<td>Numeric - Nominal</td>
<td>No = 0, ESL Program = 1, Parental Denial of the Dual Lang, bilingual, and ESL = 2, Secondary student enrolled in English courses = 3, Served in the Newcomer Program = 4, Served in the One-Way Dual Lang Prog = 5, Served in the Two-Way Dual Language Program = 6</td>
</tr>
<tr>
<td>Primary Language</td>
<td>Numeric - Nominal</td>
<td>English = 11, English/Spanish = 12, Spanish = 41, other = from 1 to 103 except the previously coded languages</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>Numeric - Nominal</td>
<td>United States = 103, Mexico = 66, other = 1-111 except the previously coded countries</td>
</tr>
<tr>
<td>Special Education Status</td>
<td>Numeric - Nominal</td>
<td>No = 0, Yes = 1</td>
</tr>
<tr>
<td>Gifted Program Status</td>
<td>Numeric - Nominal</td>
<td>No = 0, Yes = 1</td>
</tr>
<tr>
<td>Low Socioeconomic Status</td>
<td>Numeric - Nominal</td>
<td>No = 0, Yes = 1</td>
</tr>
<tr>
<td>Free/reduced lunch Status</td>
<td>Numeric - Nominal</td>
<td>Not Eligible for Free Meals = 0, Eligible for Free Meals = 1, other disadvantages = 3</td>
</tr>
<tr>
<td>Disability Status</td>
<td>Numeric - Nominal</td>
<td>No Disability = 0, Auditory Impairment = 1, Autism = 2, Developmental Delay = 3, Emotional Disturbance = 4, Learning Disability = 5, Mental Retardation = 6, Noncategorical Early Childhood = 7, Orthopedic Impairment = 8, Other Health Impairment = 9, Speech Impairment = 10, Traumatic Brain Injury = 11, Visual Impairment = 12</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>Numeric - Nominal</td>
<td>No = 0, Yes = 1</td>
</tr>
<tr>
<td>At-risk student</td>
<td>Numeric - Nominal</td>
<td>No = 0, Yes = 1</td>
</tr>
<tr>
<td>Homelessness</td>
<td>Numeric - Nominal</td>
<td>No = 0, Yes = 1</td>
</tr>
<tr>
<td>Age in days</td>
<td>Numeric-Scale</td>
<td>2190 - 5110</td>
</tr>
<tr>
<td>Istation test score</td>
<td>Numeric-Nominal</td>
<td>Grade K (Scores range 196-293), $N = 8088$</td>
</tr>
<tr>
<td>Variable Name</td>
<td>Variable Type</td>
<td>Definition</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>------------</td>
</tr>
</tbody>
</table>
| Reading section of TerraNova (100-915) | Numeric-Scale | Grade K (Scores range 355-626), N = 7821  
| | | Grade 1 (Scores range 355-701), N = 5742  
| | | Grade 2 (Scores range 314-722), N = 6705  |
| Math section of TerraNova (100-915) | Numeric-Scale | Grade K (Scores range 290-629), N = 11175  
| | | Grade 1 (Scores range 290-680), N = 10014  
| | | Grade 2 (Scores range 324-720), N = 9668  |
| Supera-Reading (100-915) | Numeric-Scale | Grade K (Scores range 355-722), N = 3354  
| | | Grade 1 (Scores range 355-722), N = 4272  
| | | Grade 2 (Scores range 355-722), N = 2963  |
| Reading section of STAAR | Numeric-Scale | Grade 3 (Scores range 108-2065), N = 7339  
| | | Grade 4 (Scores range 100-2372), N = 4353  
| | | Grade 5 (Scores range 100-2217), N = 4925  
| | | Grade 6 (Scores range 100-2272), N = 8630  
| | | Grade 7 (Scores range 100-4339), N = 8766  
| | | Grade 8 (Scores range 105-5813), N = 7552  |
| Math section of STAAR | Numeric-Scale | Grade 3 (Scores range 118-1888), N = 7339  
| | | Grade 4 (Scores range 124-1995), N = 4353  
| | | Grade 5 (Scores range 111-2143), N = 4925  
| | | Grade 6 (Scores range 100-2143), N = 8630  
| | | Grade 7 (Scores range 101-2175), N = 8766  
| | | Grade 8 (Scores range 100-4024), N = 7552  |

However, the propensity score estimation model was re-specified for some of the grade levels that did not meet the pre-specified criteria. The evaluation of the balance checking indicates that the matching of the students who responded to school readiness (ISIP) and math (TerraNova) was effective in reducing the bias for both matching algorithms. Yet, the balance checking of the Grades K to 2 students who responded to the reading assessments TerraNova(Supera) presents a differentiated picture. Splitting the sample of HIPPY and non-HIPPY into two subgroups (TerraNova and Supera) led to an imbalance between the groups on the age in days covariate. The propensity score estimation model was then re-specified by excluding the age in days covariate, which resulted in an effective balance for the fully matched samples. The disability status covariate was also removed from TerraNova math for Grades K to 2 and Supera Grade K. HIPPY, and non-HIPPY participants in Grades 3, 4, and 7 were balanced effectively for the fully matched samples. The disability status covariate behaved oddly; it was excluded from the Grade 5 samples due to small sample size and non-overlapping between the
two groups. Age in days covariate was excluded from the Grade 8 model due to the extremely high values, which resulted in increasing the variance between the groups. Lunch status covariate was removed from Grade 6 model. The re-specified models resulted in an effective balance between the groups for either the fully or optimally matched samples.

Furthermore, visual inspection was conducted as well to check for the balance between the groups. Figures 3-7 presents the standardized differences (bias) between HIPPY and non-HIPPY participants for the covariates before and after the optimal full and the optimal matching.

The propensity scores distributions of the matched HIPPY and non-HIPPY groups were similar for all grade levels, indicating that the matching process was externally valid. Figures 8-12 shows the propensity scores distribution before and after matching for each grade level. Optimal full matching retained all participants, whereas optimal matching excluded participants from each grade level. Table 6 presents the sample size of HIPPY participants and non-HIPPY participants before and after matching.

Subsequently, the ATT was estimated for the fully matched groups through regressing the standardized test scores (ISIP, TerraNova, Supera, and STAAR) on the binary treatment variable. The resulted standardized regression coefficients are the ATT estimate. Table 7 shows the treatment effect of TX HIPPY on reading and math achievement resulted from the fully matched groups for each grade level along with the treatment effect of HIPPY on school readiness measure for Grade K. The ATT of kindergarten readiness was higher for HIPPY participates with a statistically significant average treatment effect of .68. The ATT on math achievement was statistically significantly higher for HIPPY children for Grades 3, 5, and 6 with an average treatment effect of .26, .38, and .22, respectively. Similarly, the ATT on reading achievement was statistically significantly higher for HIPPY children for Grades 3, 5, and 6 with
an average treatment effect of .20, .40, and .15, respectively. The sensitivity analysis was conducted for each assessment and grade level to assess the quality of estimated treatment effects. Wilcoxon signed rank test was conducted for a range of odd ratio values. There were no changes in the upper bond p values resulted from changing the ratio values, indicating that the estimated treatment effect is un-confounded and very robust to hidden bias.

Likewise, ATT was estimated for the optimally matched groups through conducting the independent sample $t$-test and effect size on the standardized test scores (ISIP, TerraNova, Supera, and STAAR) between HIPPY and non-HIPPY groups. Table 8 presents the ATT of the TX HIPPY program on reading and math achievement for each grade level resulted from the optimally matched sample. The ATT of HIPPY kindergarten readiness was statistically significantly higher than non-HIPPY children with an average treatment effect of .75. The ATT on math achievement was statistically significantly higher for Grades 3, 5, and 6 with an average treatment effect of .26, .31, and .24, respectively. The ATT on reading achievement was statistically significantly higher for Grades 3, 5, and 6 with an average treatment effect of .21, .44, and .20, respectively.

The sensitivity analysis was conducted for each assessment and grade level to assess the quality of estimated treatment effects. Wilcoxon signed rank test was conducted for a range of odd ratio values. There were no changes in the upper bond p values resulted from changing the ratio values, indicating that the estimated treatment effect is un-confounded and very robust to hidden bias. Finally, the independent sample $t$-test on the assessments’ z-scores across the grades was conducted for each assessment separately. The results show that the HIPPY participants outperformed the non-HIPPY participants on ISIP and STAAR assessments.
Table 5
Covariate Balance Checking Between the Two Groups on Each of the Balancing Variables

<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>Before Matching</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kindergarten</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.04</td>
<td>.06</td>
<td>.92</td>
<td>1.24</td>
<td>.06</td>
<td>.06</td>
<td>.00</td>
<td>1.00</td>
<td>99.75</td>
<td>.06</td>
<td>.06</td>
<td>.00</td>
</tr>
<tr>
<td>Race</td>
<td>2.97</td>
<td>2.89</td>
<td>-.18</td>
<td>7.51</td>
<td>2.94</td>
<td>2.89</td>
<td>-.11</td>
<td>1.00</td>
<td>40.49</td>
<td>2.92</td>
<td>2.89</td>
<td>-.06</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.70</td>
<td>.92</td>
<td>.22</td>
<td></td>
<td>.94</td>
<td>.92</td>
<td>-.02</td>
<td>91.61</td>
<td></td>
<td>.94</td>
<td>.92</td>
<td>-.02</td>
</tr>
<tr>
<td>Sex</td>
<td>.51</td>
<td>.51</td>
<td>.01</td>
<td></td>
<td>.52</td>
<td>.51</td>
<td>-.01</td>
<td>43.4</td>
<td></td>
<td>.53</td>
<td>.51</td>
<td>-.02</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.45</td>
<td>4.11</td>
<td>.82</td>
<td>1.55</td>
<td>4.09</td>
<td>4.11</td>
<td>.01</td>
<td>1.05</td>
<td>98.71</td>
<td>4.17</td>
<td>4.11</td>
<td>-.03</td>
</tr>
<tr>
<td>Primary Language</td>
<td>23.61</td>
<td>34.55</td>
<td>.88</td>
<td>1.43</td>
<td>34.38</td>
<td>34.55</td>
<td>.01</td>
<td>1.00</td>
<td>98.45</td>
<td>34.25</td>
<td>34.55</td>
<td>.02</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>99.85</td>
<td>101.85</td>
<td>.30</td>
<td>3.70</td>
<td>101.6</td>
<td>101.9</td>
<td>.04</td>
<td>1.52</td>
<td>87.78</td>
<td>101.99</td>
<td>101.85</td>
<td>-.02</td>
</tr>
<tr>
<td>Special Education</td>
<td>.05</td>
<td>.07</td>
<td>.02</td>
<td></td>
<td>.09</td>
<td>.07</td>
<td>-.02</td>
<td>-20.8</td>
<td></td>
<td>.08</td>
<td>.07</td>
<td>-.01</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80</td>
<td>.83</td>
<td>.03</td>
<td></td>
<td>.86</td>
<td>.83</td>
<td>-.03</td>
<td>10.68</td>
<td></td>
<td>.86</td>
<td>.83</td>
<td>-.03</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.60</td>
<td>1.66</td>
<td>.08</td>
<td>1.13</td>
<td>1.71</td>
<td>1.66</td>
<td>-.07</td>
<td>1.15</td>
<td>10.20</td>
<td>1.72</td>
<td>1.66</td>
<td>-.08</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.35</td>
<td>.51</td>
<td>.07</td>
<td>1.54</td>
<td>.64</td>
<td>.51</td>
<td>-.06</td>
<td>1.17</td>
<td>19.01</td>
<td>.65</td>
<td>.51</td>
<td>-.07</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.54</td>
<td>.83</td>
<td>.29</td>
<td></td>
<td>.83</td>
<td>.83</td>
<td>.00</td>
<td>99.30</td>
<td></td>
<td>.83</td>
<td>.83</td>
<td>.00</td>
</tr>
<tr>
<td>Homeless</td>
<td>.02</td>
<td>.01</td>
<td>-.01</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>91.23</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Age in Days</td>
<td>2409.01</td>
<td>2414.89</td>
<td>.05</td>
<td>1.22</td>
<td>2418</td>
<td>2415</td>
<td>-.02</td>
<td>1.18</td>
<td></td>
<td>2412.0</td>
<td>2414.9</td>
<td>.03</td>
</tr>
</tbody>
</table>

TerraNova - Math Measure Balance
Kindergarten

33
<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>non-HIPPY Mean</th>
<th>HIPPY Mean</th>
<th>MD</th>
<th>VR</th>
<th>non-HIPPY Mean</th>
<th>HIPPY Mean</th>
<th>MD</th>
<th>VR</th>
<th>%imp</th>
<th>non-HIPPY Mean</th>
<th>HIPPY Mean</th>
<th>MD</th>
<th>VR</th>
<th>%imp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>.04</td>
<td>.06</td>
<td>.92</td>
<td>1.22</td>
<td>.06</td>
<td>.06</td>
<td>.00</td>
<td>1.00</td>
<td>99.74</td>
<td>.06</td>
<td>.06</td>
<td>.00</td>
<td>1.00</td>
<td>99.97</td>
</tr>
<tr>
<td>Race</td>
<td>2.97</td>
<td>2.89</td>
<td>-.18</td>
<td>7.51</td>
<td>2.94</td>
<td>2.89</td>
<td>-.12</td>
<td>1.07</td>
<td>34.36</td>
<td>2.93</td>
<td>2.89</td>
<td>-.10</td>
<td>1.67</td>
<td>64.77</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.70</td>
<td>.92</td>
<td>.22</td>
<td></td>
<td>.94</td>
<td>.92</td>
<td>-.02</td>
<td></td>
<td>90.30</td>
<td>.94</td>
<td>.92</td>
<td>-.02</td>
<td></td>
<td>92.58</td>
</tr>
<tr>
<td>Sex</td>
<td>.51</td>
<td>.51</td>
<td>.01</td>
<td></td>
<td>.54</td>
<td>.51</td>
<td>-.02</td>
<td></td>
<td>-232.9</td>
<td>.54</td>
<td>.51</td>
<td>-.02</td>
<td></td>
<td>-138.1</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.30</td>
<td>4.09</td>
<td>.85</td>
<td>1.78</td>
<td>4.07</td>
<td>4.09</td>
<td>.01</td>
<td>1.06</td>
<td>99.06</td>
<td>4.09</td>
<td>4.09</td>
<td>.00</td>
<td>1.04</td>
<td>96.67</td>
</tr>
<tr>
<td>Primary Language</td>
<td>23.61</td>
<td>34.55</td>
<td>.88</td>
<td>1.43</td>
<td>34.26</td>
<td>34.55</td>
<td>.02</td>
<td>1.01</td>
<td>97.31</td>
<td>34.19</td>
<td>34.55</td>
<td>.03</td>
<td>1.02</td>
<td>97.21</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>99.87</td>
<td>101.85</td>
<td>.30</td>
<td>3.68</td>
<td>101.9</td>
<td>101.9</td>
<td>.00</td>
<td>1.14</td>
<td>100.0</td>
<td>102.15</td>
<td>101.85</td>
<td>-.04</td>
<td>1.27</td>
<td>93.21</td>
</tr>
<tr>
<td>Special Education</td>
<td>.05</td>
<td>.07</td>
<td>.02</td>
<td></td>
<td>.09</td>
<td>.07</td>
<td>-.02</td>
<td></td>
<td>-14.9</td>
<td>.08</td>
<td>.07</td>
<td>-.02</td>
<td></td>
<td>42.79</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80</td>
<td>.83</td>
<td>.03</td>
<td></td>
<td>.86</td>
<td>.83</td>
<td>-.03</td>
<td></td>
<td>4.31</td>
<td>.86</td>
<td>.83</td>
<td>-.03</td>
<td></td>
<td>3.87</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.60</td>
<td>1.66</td>
<td>.08</td>
<td>1.13</td>
<td>1.72</td>
<td>1.66</td>
<td>-.08</td>
<td>1.17</td>
<td>2.25</td>
<td>1.72</td>
<td>1.66</td>
<td>-.08</td>
<td>1.17</td>
<td>4.43</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.35</td>
<td>.51</td>
<td>.07</td>
<td>1.54</td>
<td>.62</td>
<td>.51</td>
<td>-.05</td>
<td>1.14</td>
<td>29.37</td>
<td>.66</td>
<td>.51</td>
<td>-.07</td>
<td>1.28</td>
<td>7.92</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.54</td>
<td>.83</td>
<td>.29</td>
<td></td>
<td>.83</td>
<td>.83</td>
<td>.00</td>
<td></td>
<td>99.47</td>
<td>.83</td>
<td>.83</td>
<td>.00</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Homeless</td>
<td>.02</td>
<td>.01</td>
<td>-.01</td>
<td></td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td></td>
<td>82.66</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td></td>
<td>73.54</td>
</tr>
<tr>
<td>Age in Days</td>
<td>2248.01</td>
<td>2253.89</td>
<td>.05</td>
<td>1.22</td>
<td>2255</td>
<td>2254</td>
<td>-.01</td>
<td>1.23</td>
<td>79.45</td>
<td>2252.2</td>
<td>2254</td>
<td>-.02</td>
<td>1.02</td>
<td>50.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 1</th>
<th>non-HIPPY Mean</th>
<th>HIPPY Mean</th>
<th>MD</th>
<th>VR</th>
<th>non-HIPPY Mean</th>
<th>HIPPY Mean</th>
<th>MD</th>
<th>VR</th>
<th>%imp</th>
<th>non-HIPPY Mean</th>
<th>HIPPY Mean</th>
<th>MD</th>
<th>VR</th>
<th>%imp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>.03</td>
<td>.05</td>
<td>.80</td>
<td>1.26</td>
<td>.05</td>
<td>.05</td>
<td>.01</td>
<td>1.13</td>
<td>98.65</td>
<td>.05</td>
<td>.05</td>
<td>.01</td>
<td>1.10</td>
<td>98.84</td>
</tr>
<tr>
<td>Race</td>
<td>2.96</td>
<td>2.92</td>
<td>-.11</td>
<td>9.61</td>
<td>2.94</td>
<td>2.92</td>
<td>-.08</td>
<td>1.26</td>
<td>31.28</td>
<td>2.93</td>
<td>2.92</td>
<td>-.04</td>
<td>1.10</td>
<td>66.64</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.70</td>
<td>.93</td>
<td>.23</td>
<td></td>
<td>.94</td>
<td>.93</td>
<td>-.01</td>
<td></td>
<td>95.70</td>
<td>.94</td>
<td>.93</td>
<td>-.01</td>
<td></td>
<td>94.40</td>
</tr>
<tr>
<td>Sex</td>
<td>.51</td>
<td>.50</td>
<td>-.01</td>
<td></td>
<td>.50</td>
<td>.50</td>
<td>.00</td>
<td></td>
<td>64.35</td>
<td>.45</td>
<td>.50</td>
<td>.05</td>
<td></td>
<td>-557.3</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.42</td>
<td>4.15</td>
<td>.85</td>
<td>1.52</td>
<td>4.25</td>
<td>4.15</td>
<td>-.05</td>
<td>1.14</td>
<td>94.03</td>
<td>4.37</td>
<td>4.15</td>
<td>-.11</td>
<td>1.29</td>
<td>87.36</td>
</tr>
<tr>
<td>Primary Language</td>
<td>23.71</td>
<td>35.10</td>
<td>.96</td>
<td>1.55</td>
<td>35.26</td>
<td>35.10</td>
<td>-.01</td>
<td>1.02</td>
<td>98.54</td>
<td>35.47</td>
<td>35.10</td>
<td>-.03</td>
<td>1.05</td>
<td>96.73</td>
</tr>
<tr>
<td>Grade Level/ Covariates</td>
<td>Before Matching</td>
<td></td>
<td></td>
<td></td>
<td>Optimal Full Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Optimal Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------</td>
<td>--------</td>
<td>--------</td>
<td>------------------------</td>
<td>--------</td>
<td>----------------</td>
<td>--------</td>
<td>--------</td>
<td>-----------------</td>
<td>--------</td>
<td>----------------</td>
<td>--------</td>
<td>--------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.36</td>
<td>100.74</td>
<td>.04</td>
<td>1.57</td>
<td>100.7</td>
<td>100.7</td>
<td>.00</td>
<td>1.19</td>
<td>97.44</td>
<td>101.11</td>
<td>100.74</td>
<td>-.04</td>
<td>1.14</td>
<td>1.02</td>
</tr>
<tr>
<td>Special Education</td>
<td>.06</td>
<td>.06</td>
<td>-.01</td>
<td></td>
<td>.07</td>
<td>.06</td>
<td>-.01</td>
<td></td>
<td>-145.3</td>
<td>.04</td>
<td>.06</td>
<td>.01</td>
<td></td>
<td>-144.6</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.10</td>
<td>.16</td>
<td>.06</td>
<td></td>
<td>.14</td>
<td>.16</td>
<td>.02</td>
<td></td>
<td>62.83</td>
<td>.13</td>
<td>.16</td>
<td>.03</td>
<td></td>
<td>54.26</td>
</tr>
<tr>
<td>SES Status</td>
<td>.81</td>
<td>.89</td>
<td>.09</td>
<td></td>
<td>.86</td>
<td>.89</td>
<td>.04</td>
<td></td>
<td>57.77</td>
<td>.87</td>
<td>.89</td>
<td>.03</td>
<td></td>
<td>69.84</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.61</td>
<td>1.78</td>
<td>.28</td>
<td>1.60</td>
<td>1.71</td>
<td>1.78</td>
<td>.12</td>
<td>1.27</td>
<td>58.42</td>
<td>1.73</td>
<td>1.78</td>
<td>.08</td>
<td>1.20</td>
<td>70.39</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.45</td>
<td>.49</td>
<td>.02</td>
<td>1.12</td>
<td>.69</td>
<td>.49</td>
<td>-.10</td>
<td>1.47</td>
<td>-402.9</td>
<td>.46</td>
<td>.49</td>
<td>.02</td>
<td>1.01</td>
<td>1.41</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td></td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td></td>
<td>17.26</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td></td>
<td>-60.94</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.54</td>
<td>.83</td>
<td>.30</td>
<td></td>
<td>.86</td>
<td>.83</td>
<td>-.02</td>
<td></td>
<td>91.67</td>
<td>.89</td>
<td>.83</td>
<td>-.06</td>
<td></td>
<td>80.46</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.02</td>
<td>.00</td>
<td></td>
<td>.03</td>
<td>.02</td>
<td>-.01</td>
<td></td>
<td>-47.78</td>
<td>.02</td>
<td>.02</td>
<td>.00</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Age in Days</td>
<td>2629.56</td>
<td>2615.27</td>
<td>-.12</td>
<td>1.39</td>
<td>2614</td>
<td>2615</td>
<td>.01</td>
<td>1.18</td>
<td>92.76</td>
<td>2623.2</td>
<td>2615.3</td>
<td>-.07</td>
<td>1.36</td>
<td>44.29</td>
</tr>
<tr>
<td>Distance</td>
<td>.03</td>
<td>.04</td>
<td>.71</td>
<td>1.08</td>
<td>.04</td>
<td>.04</td>
<td>.00</td>
<td>1.04</td>
<td>99.49</td>
<td>.04</td>
<td>.04</td>
<td>.00</td>
<td>1.01</td>
<td>99.87</td>
</tr>
<tr>
<td>Race</td>
<td>2.96</td>
<td>2.89</td>
<td>-.17</td>
<td>6.21</td>
<td>2.89</td>
<td>2.89</td>
<td>-.01</td>
<td>1.35</td>
<td>94.18</td>
<td>2.89</td>
<td>2.89</td>
<td>.01</td>
<td>1.41</td>
<td>92.40</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.74</td>
<td>.87</td>
<td>.13</td>
<td></td>
<td>.88</td>
<td>.87</td>
<td>-.01</td>
<td></td>
<td>88.97</td>
<td>.89</td>
<td>.87</td>
<td>-.02</td>
<td></td>
<td>85.09</td>
</tr>
<tr>
<td>Sex</td>
<td>.52</td>
<td>.50</td>
<td>-.02</td>
<td></td>
<td>.45</td>
<td>.50</td>
<td>.04</td>
<td></td>
<td>-73.2</td>
<td>.47</td>
<td>.50</td>
<td>.03</td>
<td></td>
<td>-29.50</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.51</td>
<td>3.72</td>
<td>.55</td>
<td>1.27</td>
<td>3.76</td>
<td>3.72</td>
<td>-.02</td>
<td>1.04</td>
<td>96.75</td>
<td>3.81</td>
<td>3.72</td>
<td>-.04</td>
<td>1.06</td>
<td>92.97</td>
</tr>
<tr>
<td>Primary Language</td>
<td>24.60</td>
<td>33.40</td>
<td>.68</td>
<td>1.31</td>
<td>33.42</td>
<td>33.40</td>
<td>.00</td>
<td>1.01</td>
<td>99.74</td>
<td>33.63</td>
<td>33.40</td>
<td>-.02</td>
<td>1.02</td>
<td>97.44</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.31</td>
<td>102.08</td>
<td>.30</td>
<td>3.96</td>
<td>102.2</td>
<td>102.8</td>
<td>-.02</td>
<td>1.06</td>
<td>93.05</td>
<td>101.97</td>
<td>102.08</td>
<td>.02</td>
<td>1.45</td>
<td>93.76</td>
</tr>
<tr>
<td>Special Education</td>
<td>.09</td>
<td>.08</td>
<td>-.01</td>
<td></td>
<td>.05</td>
<td>.08</td>
<td>.02</td>
<td></td>
<td>-49.2</td>
<td>.06</td>
<td>.08</td>
<td>.02</td>
<td></td>
<td>-42.44</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.13</td>
<td>.19</td>
<td>.06</td>
<td></td>
<td>.20</td>
<td>.19</td>
<td>-.02</td>
<td></td>
<td>68.66</td>
<td>.20</td>
<td>.19</td>
<td>-.02</td>
<td></td>
<td>72.73</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80</td>
<td>.87</td>
<td>.07</td>
<td></td>
<td>.84</td>
<td>.87</td>
<td>.02</td>
<td></td>
<td>70.08</td>
<td>.85</td>
<td>.87</td>
<td>.02</td>
<td></td>
<td>71.20</td>
</tr>
<tr>
<td>Grade Level/ Covariates</td>
<td>Before Matching</td>
<td>Optimal Full Matching</td>
<td>Optimal Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------------</td>
<td>-----------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.56</td>
<td>.21</td>
<td>1.36</td>
<td>1.65</td>
<td>.07</td>
<td>1.12</td>
<td>66.59</td>
<td>1.66</td>
<td>.06</td>
<td>1.11</td>
<td>72.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability Status</td>
<td>.62</td>
<td>.00</td>
<td>1.12</td>
<td>.47</td>
<td>.06</td>
<td>1.37</td>
<td>-2404</td>
<td>.50</td>
<td>.05</td>
<td>1.27</td>
<td>-2039.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.01</td>
<td>-.01</td>
<td>1.12</td>
<td>.00</td>
<td>.00</td>
<td>67.36</td>
<td></td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>41.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.56</td>
<td>.21</td>
<td>1.12</td>
<td>.77</td>
<td>.00</td>
<td>97.96</td>
<td></td>
<td>.78</td>
<td>-.01</td>
<td>97.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homelessness</td>
<td>.01</td>
<td>.00</td>
<td></td>
<td>.01</td>
<td>.00</td>
<td>53.10</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>-.01</td>
<td>-99.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in Days</td>
<td>3007.06</td>
<td>-27</td>
<td>1.46</td>
<td>2983</td>
<td>-07</td>
<td>1.13</td>
<td>75.67</td>
<td>2985.1</td>
<td>-09</td>
<td>1.23</td>
<td>68.36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TerraNova-Reading Measure Balance</th>
<th>Kindergarten</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>.04</td>
</tr>
<tr>
<td>Race</td>
<td>3.00</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.71</td>
</tr>
<tr>
<td>Sex</td>
<td>.50</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.46</td>
</tr>
<tr>
<td>Primary Language</td>
<td>23.67</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>99.82</td>
</tr>
<tr>
<td>Special Education</td>
<td>.05</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.59</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.68</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.54</td>
</tr>
<tr>
<td>Homeless</td>
<td>.02</td>
</tr>
<tr>
<td>Grade Level/ Covariates</td>
<td>Before Matching</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
</tr>
<tr>
<td><strong>Grade 1</strong></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.03</td>
</tr>
<tr>
<td>Race</td>
<td>2.95</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.70</td>
</tr>
<tr>
<td>Sex</td>
<td>.51</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.43</td>
</tr>
<tr>
<td>Primary Language</td>
<td>24.06</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.18</td>
</tr>
<tr>
<td>Special Education</td>
<td>.06</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.10</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.60</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.45</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.00</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.54</td>
</tr>
<tr>
<td>Homeless</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Grade 2</strong></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.03</td>
</tr>
<tr>
<td>Race</td>
<td>2.99</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.73</td>
</tr>
<tr>
<td>Sex</td>
<td>.51</td>
</tr>
</tbody>
</table>

37
<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>Before Matching</th>
<th>Optimal Full Matching</th>
<th>Optimal Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.46</td>
<td>3.68</td>
<td>.54</td>
</tr>
<tr>
<td>Primary Language</td>
<td>24.49</td>
<td>33.61</td>
<td>.70</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.18</td>
<td>101.41</td>
<td>.16</td>
</tr>
<tr>
<td>Special Education</td>
<td>.09</td>
<td>.08</td>
<td>-.01</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.13</td>
<td>.17</td>
<td>.04</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80</td>
<td>.86</td>
<td>.07</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.55</td>
<td>1.68</td>
<td>.17</td>
</tr>
<tr>
<td>Disability Status</td>
<td>1.12</td>
<td>-.27</td>
<td>-.38</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.54</td>
<td>.78</td>
<td>.24</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
</tr>
</tbody>
</table>

**Supera Measure Balance**

<table>
<thead>
<tr>
<th>Kindergarten</th>
</tr>
</thead>
</table>

| Distance                                | .04 | .06 | .85 | 1.07 | .06 | .06 | .00 | 1.03 | 97.80 | .06 | .06 | .00 | 1.00 | 99.83 |
| Race                                   | 2.94 | 2.92 | -.09 | 12.90 | 2.93 | 2.92 | -.05 | 1.16 | 11.46 | 2.92 | 2.92 | -.02 | 1.08 | -198.6 |
| Ethnicity                              | .70 | .92 | .22 |       | .93 | .92 | -.01 |       | 86.86 | .93 | .92 | -.01 |       | 72.84 |
| Sex                                    | .51 | .50 | .00 |       | .54 | .50 | -.03 |       | -38.5 | .53 | .50 | -.03 |       | -310.5 |
| ESL Program Status                     | 2.45 | 4.05 | .77 | 1.44 | 4.17 | 4.05 | -.05 | 1.08 | 90.09 | 4.20 | 4.05 | -.07 | 1.11 | 77.30 |
| Primary Language                       | 23.57 | 34.19 | .85 | 1.39 | 33.99 | 34.19 | .02 | 1.01 | 80.54 | 34.06 | 34.19 | .01 | 1.00 | 65.65 |
| Country of Birth                       | 99.88 | 101.53 | .21 | 2.82 | 101.89 | 101.53 | -.05 | 1.18 | 88.63 | 101.49 | 101.53 | .01 | 1.35 | 81.52 |
| Special Education                      | .05 | .08 | .04 |       | .08 | .08 | .01 |       | 67.36 | .07 | .08 | .01 |       | 31.29 |

38
<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>Before Matching</th>
<th>Optimal Full Matching</th>
<th>Optimal Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HIPPY Mean</td>
<td>HIPPY Mean MD VR</td>
<td>HIPPY Mean MD VR</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80</td>
<td>.81 .01 .01</td>
<td>.83 .81 .02 .02</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.60</td>
<td>1.62 .02 1.03</td>
<td>1.66 1.62 .06 1.10</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.54</td>
<td>.80 .26</td>
<td>.82 .80 .02 .02</td>
</tr>
<tr>
<td>Homeless</td>
<td>.02</td>
<td>.00 - .02</td>
<td>.00 .00 .00 .00</td>
</tr>
<tr>
<td>Grade 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HIPPY Mean</td>
<td>HIPPY Mean MD VR</td>
<td>HIPPY Mean MD VR</td>
</tr>
<tr>
<td>Distance</td>
<td>.02</td>
<td>.05 .71 2.07</td>
<td>.05 .05 .03 1.39</td>
</tr>
<tr>
<td>Race</td>
<td>2.96</td>
<td>2.94 -.11 18.22</td>
<td>2.95 2.94 -.07 2.12</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.70</td>
<td>.94 .23</td>
<td>.94 .94 .01 94.32</td>
</tr>
<tr>
<td>Sex</td>
<td>.50</td>
<td>.45 -.05</td>
<td>.46 .45 .00 25.6</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.42</td>
<td>4.17 .86 1.53</td>
<td>4.02 4.17 .07 1.10</td>
</tr>
<tr>
<td>Primary Language</td>
<td>23.40</td>
<td>35.22 1.00 1.56</td>
<td>34.19 35.22 .09 1.13</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.53</td>
<td>100.82 .03 1.37</td>
<td>100.15 100.82 .07 1.31</td>
</tr>
<tr>
<td>Special Education</td>
<td>.06</td>
<td>.06 .00</td>
<td>.11 .06 -.04 -900.2</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.11</td>
<td>.17 .06</td>
<td>.17 .17 .00 94.25</td>
</tr>
<tr>
<td>SES Status</td>
<td>.81</td>
<td>.88 .07</td>
<td>.88 .88 .00 -23.9</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.62</td>
<td>1.77 .23 1.46</td>
<td>1.76 1.77 .01 1.01</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.46</td>
<td>.42 -.02 1.12</td>
<td>.61 .42 -.09 1.37</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.00</td>
<td>.01 .01</td>
<td>.03 .01 -.02 -112.6</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.53</td>
<td>.82 .29</td>
<td>.80 .82 .02 90.54</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.00 -.01</td>
<td>.00 .00 .00 47.92</td>
</tr>
<tr>
<td>Grade 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HIPPY Mean</td>
<td>HIPPY Mean MD VR</td>
<td>HIPPY Mean MD VR</td>
</tr>
<tr>
<td>Distance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESL Program Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country of Birth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lunch Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyslexia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At-risk Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homeless</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

39
<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>Before Matching</th>
<th>Optimal Full Matching</th>
<th>Optimal Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean MD VR</td>
<td>non-HIPPY Mean</td>
</tr>
<tr>
<td>Distance</td>
<td>.03 .05 .75 1.02</td>
<td>.05 .05 -.02 1.16 99.35</td>
<td>.05 .05 .00 1.00</td>
</tr>
<tr>
<td>Race</td>
<td>2.92 2.87 -.15 7.28</td>
<td>2.86 2.87 .04 1.92 64.88</td>
<td>2.84 2.87 .08 1.34</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.75 .87 .12</td>
<td>.87 .87 .00 81.93</td>
<td>.85 .87 .02 54.35</td>
</tr>
<tr>
<td>Sex</td>
<td>.53 .53 .00</td>
<td>.56 .53 -.02 -383.5</td>
<td>.53 .53 .00 -281.6</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.56 3.76 .55 1.31</td>
<td>3.86 3.76 -.05 1.02 76.46</td>
<td>3.86 3.76 -.05 1.02 48.42</td>
</tr>
<tr>
<td>Primary Language</td>
<td>24.71 33.22 .65 1.29</td>
<td>33.13 33.22 .01 1.04 77.32</td>
<td>33.36 33.22 -.01 1.02 52.54</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.45 102.65 .62 10.06</td>
<td>102.84 102.65 -.05 1.48 92.61</td>
<td>102.48 102.65 .05 1.43 82.17</td>
</tr>
<tr>
<td>Special Education</td>
<td>.09 .07 -.02</td>
<td>.08 .07 -.01 -25.7</td>
<td>.07 .07 .00 -138.1</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.13 .20 .07</td>
<td>.17 .20 .02 67.39</td>
<td>.16 .20 .04 86.66</td>
</tr>
<tr>
<td>SES Status</td>
<td>.80 .87 .07</td>
<td>.86 .87 .01 61.64</td>
<td>.88 .87 -.01 61.60</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.56 1.72 .24 1.40</td>
<td>1.71 1.72 .02 1.03 70.25</td>
<td>1.74 1.72 -.03 1.06 77.12</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.65 .70 .02 1.26</td>
<td>.71 .70 .00 1.03 -224.7</td>
<td>.69 .70 .00 1.03 -417.5</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.01 .00 -.01</td>
<td>.00 .00 .00 94.47</td>
<td>.00 .00 .00 100.00</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.58 .76 .18</td>
<td>.78 .76 -.02 72.52</td>
<td>.77 .76 -.01 32.38</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01 .00 -.01</td>
<td>.00 .00 .00 94.47</td>
<td>.00 .00 .00 100.00</td>
</tr>
<tr>
<td><strong>STAAR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grade 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.02 .05 .90 1.34</td>
<td>.05 .05 .01 1.09 98.96</td>
<td>.05 .05 .01 1.10 98.84</td>
</tr>
<tr>
<td>Race</td>
<td>2.89 2.95 .12 3.06</td>
<td>2.95 2.95 .00 1.79 99.25</td>
<td>2.99 2.95 -.07 2.44 45.96</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.72 .90 .18</td>
<td>.89 .90 .00 99.50</td>
<td>.87 .90 .02 86.76</td>
</tr>
<tr>
<td>Sex</td>
<td>.51 .53 .02</td>
<td>.55 .53 -.03 -28.1</td>
<td>.54 .53 -.01 40.83</td>
</tr>
</tbody>
</table>

40
<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>Before Matching</th>
<th></th>
<th></th>
<th></th>
<th>Optimal Full Matching</th>
<th></th>
<th></th>
<th>%imp</th>
<th>Optimal Matching</th>
<th></th>
<th></th>
<th>%imp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.43</td>
<td>3.90</td>
<td>.70</td>
<td>1.39</td>
<td>3.83</td>
<td>3.90</td>
<td>.03</td>
<td>1.01</td>
<td>95.80</td>
<td>3.89</td>
<td>3.90</td>
<td>.00</td>
</tr>
<tr>
<td>Primary Language</td>
<td>24.73</td>
<td>34.95</td>
<td>.85</td>
<td>1.55</td>
<td>35.09</td>
<td>34.95</td>
<td>-.01</td>
<td>1.00</td>
<td>98.62</td>
<td>35.32</td>
<td>34.95</td>
<td>-.03</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.13</td>
<td>102.35</td>
<td>.46</td>
<td>7.05</td>
<td>101.9</td>
<td>102.4</td>
<td>.10</td>
<td>2.28</td>
<td>78.96</td>
<td>101.37</td>
<td>102.35</td>
<td>.20</td>
</tr>
<tr>
<td>Special Education</td>
<td>.08</td>
<td>.05</td>
<td>-.03</td>
<td></td>
<td>.06</td>
<td>.05</td>
<td>-.01</td>
<td>58.92</td>
<td></td>
<td>.05</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.14</td>
<td>.23</td>
<td>.08</td>
<td></td>
<td>.24</td>
<td>.23</td>
<td>-.01</td>
<td>88.38</td>
<td></td>
<td>.24</td>
<td>.23</td>
<td>-.01</td>
</tr>
<tr>
<td>SES Status</td>
<td>.90</td>
<td>.98</td>
<td>.08</td>
<td></td>
<td>.98</td>
<td>.98</td>
<td>.00</td>
<td>99.09</td>
<td></td>
<td>.98</td>
<td>.98</td>
<td>-.01</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.32</td>
<td>1.12</td>
<td>-.50</td>
<td>2.72</td>
<td>1.13</td>
<td>1.12</td>
<td>-.02</td>
<td>1.03</td>
<td>95.42</td>
<td>1.13</td>
<td>1.12</td>
<td>-.03</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.59</td>
<td>.16</td>
<td>-.46</td>
<td>4.95</td>
<td>.19</td>
<td>.16</td>
<td>-.03</td>
<td>1.03</td>
<td>92.42</td>
<td>.13</td>
<td>.16</td>
<td>.03</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.03</td>
<td>.01</td>
<td>-.02</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>-.01</td>
<td>54.62</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>-.01</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.55</td>
<td>.78</td>
<td>.23</td>
<td></td>
<td>.79</td>
<td>.78</td>
<td>-.01</td>
<td>94.52</td>
<td></td>
<td>.80</td>
<td>.78</td>
<td>-.02</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>66.91</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Age in Days</td>
<td>3387.62</td>
<td>3358.84</td>
<td>-.22</td>
<td>1.49</td>
<td>3356</td>
<td>3359</td>
<td>.03</td>
<td>1.10</td>
<td>88.68</td>
<td>3362.5</td>
<td>3358.8</td>
<td>-.03</td>
</tr>
<tr>
<td></td>
<td>Grade 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.02</td>
<td>.03</td>
<td>.78</td>
<td>1.10</td>
<td>.03</td>
<td>.03</td>
<td>-.02</td>
<td>1.17</td>
<td>97.89</td>
<td>.03</td>
<td>.03</td>
<td>-.01</td>
</tr>
<tr>
<td>Race</td>
<td>2.91</td>
<td>2.92</td>
<td>.01</td>
<td>2.60</td>
<td>2.94</td>
<td>2.92</td>
<td>-.04</td>
<td>1.52</td>
<td>-160.8</td>
<td>2.89</td>
<td>2.92</td>
<td>.04</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.75</td>
<td>.86</td>
<td>.11</td>
<td></td>
<td>.85</td>
<td>.86</td>
<td>.01</td>
<td>92.19</td>
<td></td>
<td>.85</td>
<td>.86</td>
<td>.01</td>
</tr>
<tr>
<td>Sex</td>
<td>.50</td>
<td>.49</td>
<td>-.01</td>
<td></td>
<td>.45</td>
<td>.49</td>
<td>.04</td>
<td>-522.1</td>
<td></td>
<td>.45</td>
<td>.49</td>
<td>.05</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.50</td>
<td>3.74</td>
<td>.56</td>
<td>1.24</td>
<td>3.92</td>
<td>3.74</td>
<td>-.08</td>
<td>1.14</td>
<td>85.31</td>
<td>3.85</td>
<td>3.74</td>
<td>-.05</td>
</tr>
<tr>
<td>Primary Language</td>
<td>25.68</td>
<td>33.31</td>
<td>.58</td>
<td>1.31</td>
<td>33.50</td>
<td>33.31</td>
<td>-.01</td>
<td>1.01</td>
<td>97.48</td>
<td>33.41</td>
<td>33.31</td>
<td>-.01</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>99.64</td>
<td>102.56</td>
<td>.73</td>
<td>12.66</td>
<td>102.64</td>
<td>102.56</td>
<td>-.02</td>
<td>1.25</td>
<td>97.35</td>
<td>103.13</td>
<td>102.56</td>
<td>-.14</td>
</tr>
<tr>
<td>Special Education</td>
<td>.09</td>
<td>.07</td>
<td>-.02</td>
<td></td>
<td>.08</td>
<td>.07</td>
<td>-.01</td>
<td>42.37</td>
<td></td>
<td>.07</td>
<td>.07</td>
<td>.00</td>
</tr>
</tbody>
</table>

<p>| 41 |</p>
<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>Before Matching</th>
<th>Optimal Full Matching</th>
<th>Optimal Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>non-HIPPY Mean</td>
<td>non-HIPPY Mean</td>
</tr>
<tr>
<td></td>
<td>HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>HIPPY Mean</td>
</tr>
<tr>
<td></td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.20</td>
<td>.25</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>.27</td>
<td>.27</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>.07</td>
<td>.03</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES Status</td>
<td>.93</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>.96</td>
<td>.96</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>.03</td>
<td>-.01</td>
<td>-.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.17</td>
<td>1.03</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>-.55</td>
<td>-.08</td>
<td>-.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability Status</td>
<td>.64</td>
<td>.83</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>-.03</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.04</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>-.03</td>
<td>-.01</td>
<td>-.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.73</td>
<td>.89</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>.85</td>
<td>.85</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>.12</td>
<td>-.05</td>
<td>-.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>-.01</td>
<td>-.01</td>
<td>-.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in Days</td>
<td>3762.99</td>
<td>3742</td>
<td>3754.4</td>
</tr>
<tr>
<td></td>
<td>3727.41</td>
<td>-.05</td>
<td>-.11</td>
</tr>
<tr>
<td></td>
<td>-28.0</td>
<td>1.12</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>1.68</td>
<td>59.21</td>
<td>24.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.02</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>.03</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.84</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>1.29</td>
<td>99.99</td>
<td>99.71</td>
</tr>
<tr>
<td>Race</td>
<td>2.95</td>
<td>2.99</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>2.95</td>
<td>.02</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>.02</td>
<td>3.30</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-435.2</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.74</td>
<td>.87</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>.90</td>
<td>.90</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td>.15</td>
<td>.03</td>
<td>.09</td>
</tr>
<tr>
<td>Sex</td>
<td>.52</td>
<td>.55</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>.49</td>
<td>.49</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>-.03</td>
<td>-.06</td>
<td>-.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-38.10</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>2.40</td>
<td>3.75</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td>3.89</td>
<td>3.89</td>
<td>3.89</td>
</tr>
<tr>
<td></td>
<td>.72</td>
<td>.72</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>1.41</td>
<td>91.21</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>82.17</td>
</tr>
<tr>
<td>Primary Language</td>
<td>25.35</td>
<td>33.53</td>
<td>32.97</td>
</tr>
<tr>
<td></td>
<td>34.49</td>
<td>34.49</td>
<td>34.49</td>
</tr>
<tr>
<td></td>
<td>.74</td>
<td>.74</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td>1.46</td>
<td>89.40</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>83.28</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>10.67</td>
<td>102.9</td>
<td>103.06</td>
</tr>
<tr>
<td></td>
<td>102.99</td>
<td>102.9</td>
<td>102.99</td>
</tr>
<tr>
<td></td>
<td>21.7</td>
<td>.79</td>
<td>-.64</td>
</tr>
<tr>
<td></td>
<td>101.8</td>
<td>490.2</td>
<td>37.14</td>
</tr>
<tr>
<td>Special Education</td>
<td>.09</td>
<td>.08</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>-.05</td>
<td>-.03</td>
<td>-.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75.46</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.21</td>
<td>.42</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>.38</td>
<td>.38</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td>.17</td>
<td>-.04</td>
<td>-.01</td>
</tr>
<tr>
<td>SES Status</td>
<td>.92</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>.08</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.16</td>
<td>1.05</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>-.36</td>
<td>.06</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>4.44</td>
<td>1.14</td>
<td>1.93</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.62</td>
<td>.68</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>.36</td>
<td>.36</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>-.16</td>
<td>-.09</td>
<td>-.03</td>
</tr>
<tr>
<td></td>
<td>1.42</td>
<td>1.68</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-24.3</td>
<td>77.92</td>
</tr>
<tr>
<td>Grade Level/ Covariates</td>
<td>Before Matching</td>
<td>Optimal Full Matching</td>
<td>Optimal Matching</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------</td>
<td>-----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td>non-HIPPY</td>
<td>HIPPY</td>
<td>non-HIPPY</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.03</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.77</td>
<td>.90</td>
<td>.87</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Age in Days</td>
<td>4099.02</td>
<td>4090.45</td>
<td>4101</td>
</tr>
<tr>
<td>Grade 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Race</td>
<td>2.94</td>
<td>2.92</td>
<td>2.90</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.72</td>
<td>.92</td>
<td>.90</td>
</tr>
<tr>
<td>Sex</td>
<td>.52</td>
<td>.45</td>
<td>.43</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>1.45</td>
<td>2.23</td>
<td>1.99</td>
</tr>
<tr>
<td>Primary Language</td>
<td>24.94</td>
<td>33.56</td>
<td>33.26</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>100.02</td>
<td>103.00</td>
<td>102.97</td>
</tr>
<tr>
<td>Special Education</td>
<td>.10</td>
<td>.08</td>
<td>.09</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.19</td>
<td>.39</td>
<td>.39</td>
</tr>
<tr>
<td>SES Status</td>
<td>.85</td>
<td>.93</td>
<td>.94</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.05</td>
<td>.03</td>
<td>.01</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.76</td>
<td>.83</td>
<td>.80</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Age in Days</td>
<td>4509.9</td>
<td>4491.02</td>
<td>4487.8</td>
</tr>
<tr>
<td>Grade 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Grade Level/ Covariates</td>
<td>Before Matching</td>
<td>Optimal Full Matching</td>
<td>Optimal Matching</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------</td>
<td>----------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean MD VR %imp</td>
<td>non-HIPPY Mean</td>
</tr>
<tr>
<td>Race</td>
<td>2.89 2.92 .07 4.74</td>
<td>2.96 2.92 -.09 1.39 -40.7</td>
<td>2.97 2.92 -.13 3.71 -102.6</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.73  .95  .22</td>
<td>.97  .95  -.02 1.39  92.97</td>
<td>.98  .95  -.03 87.65</td>
</tr>
<tr>
<td>Sex</td>
<td>.50  .52  .02</td>
<td>.51  .52  .02 1.06  97.12</td>
<td>.53  .52  -.01 58.86</td>
</tr>
<tr>
<td>ESL Program Status</td>
<td>1.42 2.45 .56 1.19</td>
<td>2.48 2.45 .02 1.06  97.12</td>
<td>2.54 2.45 -.05 1.06  91.10</td>
</tr>
<tr>
<td>Primary Language</td>
<td>25.49 36.39 .98 1.84</td>
<td>37.10 36.39 .06 1.19  93.44</td>
<td>36.63 36.39 -.02 1.09  97.73</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>99.22 101.62 .34 3.91</td>
<td>101.5 101.6 .02 1.34  94.61</td>
<td>102.32 101.62 -.10 2.00  71.00</td>
</tr>
<tr>
<td>Special Education</td>
<td>.09  .06  -.03</td>
<td>.05  .06  .00 90.30</td>
<td>.07  .06  -.02 43.16</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.17  .28  .12</td>
<td>.29  .28  -.01 91.13</td>
<td>.28  .28  .01 92.09</td>
</tr>
<tr>
<td>SES Status</td>
<td>.85  .95  .09</td>
<td>.94  .95  .00 99.81</td>
<td>.95  .95  .00 100.00</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.02 1.02 -.01 2.47</td>
<td>1.03 1.02 -.04 1.11 -183.5</td>
<td>1.04 1.02 -.05 1.14 -251.8</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.61  .44 -.10 1.32</td>
<td>.34  .44 .06 1.24 38.12</td>
<td>.50  .44 -.03 1.18 66.76</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.04  .03  -.01</td>
<td>.01  .03 .02 -128.4</td>
<td>.00  .03 .03 -246.65</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.76  .83  .07</td>
<td>.86  .83  -.03 57.48</td>
<td>.86  .83  -.04 47.31</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01  .00  -.01</td>
<td>.00  .00 .00 96.48</td>
<td>.00  .00 .00 100.00</td>
</tr>
<tr>
<td>Age in Days</td>
<td>4881.16 4854.61 -.13 1.16</td>
<td>4865 4855  -.05 1.29 60.42</td>
<td>4871.4 4854.6  -.08 1.33 36.60</td>
</tr>
</tbody>
</table>

Grade 8

<p>| Distance                | .01  .02  .62 2.09 | .02  .02  -.01 1.08 99.53 | .02  .02  .00 1.01 97.88 |
| Race                    | 2.90 2.86 -.11 4.69 | 2.84 2.86 .03 1.47 48.92 | 2.86 2.86 .00 1.19 49.59 |
| Ethnicity               | .73  .88  .16   | .86  .88  .02 86.59 | .87  .88  .01 54.22 |
| Sex                     | .51  .58  .07   | .56  .58  .03 47.52 | .54  .58  .04 -50.33 |
| ESL Program Status      | 1.36 2.16 .48 1.11 | 1.99 2.16  .09 1.44 93.78 | 2.29 2.16 -.08 1.60 89.16 |</p>
<table>
<thead>
<tr>
<th>Grade Level/ Covariates</th>
<th>Before Matching</th>
<th></th>
<th></th>
<th></th>
<th>Optimal Full Matching</th>
<th></th>
<th></th>
<th>%imp</th>
<th>Optimal Matching</th>
<th></th>
<th></th>
<th>%imp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
<td>VR</td>
<td>%imp</td>
<td>non-HIPPY Mean</td>
<td>HIPPY Mean</td>
<td>MD</td>
</tr>
<tr>
<td>Primary Language</td>
<td>25.35</td>
<td>34.09</td>
<td>.65</td>
<td>1.23</td>
<td>33.03</td>
<td>34.09</td>
<td>.08</td>
<td>1.04</td>
<td>89.37</td>
<td>33.88</td>
<td>34.09</td>
<td>.02</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>98.74</td>
<td>100.35</td>
<td>.16</td>
<td>1.99</td>
<td>99.78</td>
<td>100.35</td>
<td>.06</td>
<td>1.86</td>
<td>62.23</td>
<td>99.86</td>
<td>100.35</td>
<td>.05</td>
</tr>
<tr>
<td>Special Education</td>
<td>.09</td>
<td>.10</td>
<td>.00</td>
<td></td>
<td>.16</td>
<td>.10</td>
<td>-.06</td>
<td></td>
<td>-2432</td>
<td>.19</td>
<td>.10</td>
<td>-.09</td>
</tr>
<tr>
<td>Talented &amp; Gifted Status</td>
<td>.16</td>
<td>.23</td>
<td>.08</td>
<td></td>
<td>.23</td>
<td>.23</td>
<td>.00</td>
<td></td>
<td>-4.22</td>
<td>.22</td>
<td>.23</td>
<td>.01</td>
</tr>
<tr>
<td>SES Status</td>
<td>.85</td>
<td>.94</td>
<td>.09</td>
<td></td>
<td>.94</td>
<td>.94</td>
<td>.00</td>
<td></td>
<td>96.58</td>
<td>.91</td>
<td>.94</td>
<td>.03</td>
</tr>
<tr>
<td>Lunch Status</td>
<td>1.02</td>
<td>.99</td>
<td>-.11</td>
<td>3.17</td>
<td>.99</td>
<td>.99</td>
<td>-.04</td>
<td>1.11</td>
<td>60.65</td>
<td>.94</td>
<td>.99</td>
<td>.14</td>
</tr>
<tr>
<td>Disability Status</td>
<td>.65</td>
<td>.41</td>
<td>-.19</td>
<td>2.33</td>
<td>.52</td>
<td>.41</td>
<td>-.09</td>
<td>1.27</td>
<td>-96.53</td>
<td>.77</td>
<td>.41</td>
<td>-.28</td>
</tr>
<tr>
<td>Dyslexia</td>
<td>.04</td>
<td>.03</td>
<td>-.01</td>
<td></td>
<td>.04</td>
<td>.03</td>
<td>-.01</td>
<td></td>
<td>-438.5</td>
<td>.07</td>
<td>.03</td>
<td>-.04</td>
</tr>
<tr>
<td>At-risk Status</td>
<td>.76</td>
<td>.83</td>
<td>.07</td>
<td></td>
<td>.84</td>
<td>.83</td>
<td>-.01</td>
<td></td>
<td>24.75</td>
<td>.87</td>
<td>.83</td>
<td>-.04</td>
</tr>
<tr>
<td>Homeless</td>
<td>.01</td>
<td>.00</td>
<td>-.01</td>
<td></td>
<td>.001</td>
<td>0</td>
<td>-.00</td>
<td></td>
<td>69.96</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note: MD = Mean Difference. VR = Variance Ratio. %imp = % improved.
The independent sample $t$-test on ISIP was found to be statistically significant,
$t(8086) = -11.42, p < .001; d = .65$, the effect size for this analysis was found to go above Cohen’s (1988) suggestion for a medium effect ($d = .50$). The independent sample $t$-test on TerraNova math, TerraNova reading, and Supera reading assessments were found to be not statistically significant. Similarly, the independent sample $t$-test was conducted on STAAR math and reading assessments, the results were found statistically significant for math with small effect size, $t(41,007) = -2.40, p = .02; d = .10$.

Discussion

TX HIPPY Program Evaluation

This evaluation estimated the impact of the TX HIPPY program on math and reading achievement using quasi-experimental design and applying optimal full and optimal propensity score matching algorithms. The intention of using the optimal full and optimal matching algorithms was to assess the potential selection bias as part of the impact evaluation of the program. The ATT estimates showed that the TX HIPPY program positively influences kindergarten students to start school ready to learn. The statistically significant average treatment effect supports the assumption of continually having a substantial and practical effect. In essence, we can state that the TX HIPPY program has a positive influence on the participants’ school readiness skills.

The evaluation findings for the school readiness measure also agree with the body of HIPPY research that supports the TX HIPPY program’s essential beliefs that children’s education starts in the home (BarHava-Monteith, Harre, & Field, 1999; Bradley & Gilkey, 2003; Brown & Johnson, 2014; Garcia, 2006; Jacobson, 2003). Nevertheless, the school readiness outcome for kindergarten is lower once children who attended the Early Learning Program, for
example, are accounted for. As well, when children do attend Pre-K, their likelihood of being school ready in kindergarten increases by 350% (DISD, 2015). Therefore, the exposure to other programs and the enrollment in Pre-K are possible confounds.

The findings also show that ISIP assessment for kindergarten children’s school readiness in the winter of 2017 were inconsistent predictors of children’s math and reading skills measured by TerraNova and Supera assessments in the spring of 2018. This finding agrees with evidence that children transitioning to formal schooling in kindergarten sit in a classroom that is more learning-oriented than social-oriented, which makes it less attractive (Hamre et al., 2013). Another possibility is that the fade-out of immediate effects might be due to the inadequate effectiveness of public schools (Barnett, 2011).

Concerning the findings of math and reading achievements, a more differentiated picture emerges. HIPPY children were found to have higher math and reading achievement; the difference reached statistical significance on three grade levels (Grades 3, 5, and 6) with a modest positive impact. This finding supports the results of Bradley and Gilkey’s (2003) evaluation study of the TX HIPPY program. However, the treatment effects of the TX HIPPY program on math and reading achievement were not statistically significant for other grade levels. This finding supports the previous research findings of BarHava-Monteith, Brown, and Johnson (2014), Harre and Field (1999), and Smith (1995); their interpretation of the non-statistically significant findings is that the treatment effect vanishes as the students progressed in school over the years. Although the results of this evaluation align with previous research results, the interpretation of the findings differs.

The TX HIPPY program targets a special population of families identified as having limited education, low-income, or lack social support who their children are at risk of facing
deficiencies regarding their intellectual abilities, emotional support, and availability of resources. These deficiencies can prevent them from starting school ready to learn and can be predictive of their academic success in the following years (Shonkoff & Phillips, 2000). Accordingly, the non-statistically significant difference in the performance between HIPPY and non-HIPPY participants could be considered as gain. The mixed findings suggest that HIPPY children scored at the same level or higher than non-HIPPY children did on math and reading achievement, indicating that TX HIPPY program has achieved its goal of helping children maintain long-term academic success.

Though the TX HIPPY program can be an important opportunity for low-income families, its quality varies, and the exposure period is uneven. The TX HIPPY program enrolls families of three or four-year-old children for at least two program years and more research is needed to determine whether the outcome differs if children enroll at age three versus age four. Home visiting research shows that families generally receive half of the visits expected by the program model with 20% to 80% dropping out before the program is scheduled to end (Gomby, Culross, & Behrman, 1999). The available data on the HIPPY program indicate that the attrition rate may reach 28% (Baker, Piotrkowski, & Brooks-Gunn, 1998). Another problem involves what families decide by the end of one-program-year; for example, many children who completed one-year-program did not enroll the following year. Hence, the mixed findings can be investigated in light of the level and intensity of HIPPY exposure in addition to the quality of the delivered service.

Moreover, the impact evaluation of an early childhood education program such as HIPPY should include two components, the parent and the child. According to Chase-Lansdale and Brooks-Gunn (2014), programs should focus on a two-generation program strategy. Early
childhood education interventions help children advance in learning and might influence their parents to get more education and better jobs. Likewise, better-educated parents may support children to become more engaged, motivated, and successful (Klebanov & Brooks-Gunn, 2006).
Figure 3. Love plots before and after propensity score matching for Grade K.
Figure 4. Love plots before and after propensity score matching for Grade 1.
Figure 5. Love plots before and after propensity score matching for Grade 2.
Figure 6. Love plots before and after propensity score matching for Grades 3, 4, and 5.
Figure 7. Love plots before and after propensity score matching for Grades 6, 7, and 8.
Figure 8. Propensity scores distribution before and after the matching for Grade K.
Figure 9. Propensity scores distribution before and after the matching for Grade 1.
Figure 10. Propensity scores distribution before and after the matching for Grade 2.
Figure 11. Propensity scores distribution before and after the matching for Grades 3, 4, and 5.
Figure 12. Propensity scores distribution before and after the matching for Grades 6, 7, and 8.
Table 6

Sample Size of HIPPY Participants and Non-HIPPY participants Before and After Matching

<table>
<thead>
<tr>
<th>Grade</th>
<th>ISIP</th>
<th>TerraNova - Math</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Optimal Full Matching</td>
<td>Optimal Matching</td>
</tr>
<tr>
<td></td>
<td>non-HIPPY</td>
<td>HIPPY</td>
</tr>
<tr>
<td></td>
<td>Kindergarten</td>
<td>Kindergarten</td>
</tr>
<tr>
<td>All</td>
<td>7780</td>
<td>308</td>
</tr>
<tr>
<td>Matched</td>
<td>7780</td>
<td>308</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 1</td>
<td>TerraNova - Reading</td>
<td>Kindergarten</td>
</tr>
<tr>
<td>All</td>
<td>5768</td>
<td>156</td>
</tr>
<tr>
<td>Matched</td>
<td>5768</td>
<td>156</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>6716</td>
<td>200</td>
</tr>
<tr>
<td>Matched</td>
<td>6716</td>
<td>200</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade</td>
<td>Optimal Full Matching</td>
<td>Optimal Matching</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td>non-HIPPY</td>
<td>HIPPY</td>
</tr>
<tr>
<td>Grade 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>2723</td>
<td>79</td>
</tr>
<tr>
<td>Matched</td>
<td>2723</td>
<td>79</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>3444</td>
<td>93</td>
</tr>
<tr>
<td>Matched</td>
<td>3444</td>
<td>93</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Supera</td>
<td>Kindergarten</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4156</td>
<td>165</td>
</tr>
<tr>
<td>Matched</td>
<td>4156</td>
<td>165</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>3045</td>
<td>77</td>
</tr>
<tr>
<td>Matched</td>
<td>3045</td>
<td>77</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>3272</td>
<td>107</td>
</tr>
<tr>
<td>Matched</td>
<td>3272</td>
<td>107</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>STAAR - Math &amp; Science</td>
<td>Grade 3</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>7167</td>
<td>172</td>
</tr>
<tr>
<td>Matched</td>
<td>7167</td>
<td>172</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade</td>
<td>Optimal Full Matching</td>
<td>Optimal Matching</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td>non-HIPPY</td>
<td>HIPPY</td>
</tr>
<tr>
<td>Grade 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4268</td>
<td>85</td>
</tr>
<tr>
<td>Matched</td>
<td>4268</td>
<td>85</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4838</td>
<td>87</td>
</tr>
<tr>
<td>Matched</td>
<td>4838</td>
<td>87</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>8534</td>
<td>96</td>
</tr>
<tr>
<td>Matched</td>
<td>8534</td>
<td>96</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>8657</td>
<td>109</td>
</tr>
<tr>
<td>Matched</td>
<td>8657</td>
<td>109</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>7483</td>
<td>69</td>
</tr>
<tr>
<td>Matched</td>
<td>7483</td>
<td>69</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The parent-child synergistic effect could consequence adverse outcomes as well; the demands of the child, school, and employment might upsurge parents’ stress and force them to spend time apart from their children (Chase-Lansdale & Brooks-Gunn, 2014). Therefore, the mixed findings can be examined by adding variables related to parents’ level of engagement in
their child’s education in the home, at school, and in the community. Those parent-child interaction level and social support system may make the evaluation more reliable.

An additional point to consider when rationalizing the mixed findings is the complex social contexts of the classroom and school that may interact in substantial ways to either limit or support children’s academic growth (Lowenstein, Friedman-Krauss, Raver, Jones, & Pess, 2015). Children’s learning is multidimensional; many factors beyond their academic skills (e.g., teacher-teacher relationship, classroom organization) contribute to their short- and long-term academic success (Hamre & Pianta, 2001; Pianta & Hamre, 2009; Pianta & Stuhlman, 2004). A close teacher-child relationship predicts socioeconomically disadvantaged children’s higher literacy and math scores and moderates the relationship between children’s academic achievement and low parent support during the transition to kindergarten (Lowenstein et al., 2015). The classrooms in low-income communities often have less experienced teachers and limited resources and facilities (Darling-Hammond, 2004). Accordingly, DISD needs to expand the enrollment in early childhood education and follow-up programs to include all the children in the district to increase the number of children who enter school ready for success and maintain that success across the grades.

Table 7

*Treatment Effect of TX HIPPY Program on Reading and Math Achievement for Each Grade Level Resulted from the Fully Matched Sample*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Outcome</th>
<th>t</th>
<th>p</th>
<th>ATT b</th>
<th>SE b</th>
<th>ATT β</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Adjuste R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISIP-Kindergarten</td>
<td>11.47</td>
<td>.00</td>
<td>166.45</td>
<td>14.51</td>
<td>.68</td>
<td>131.70</td>
<td>8,086</td>
<td>.000**</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>TerraNova-Math</td>
<td>.41</td>
<td>.68</td>
<td>6.95</td>
<td>16.85</td>
<td>.02</td>
<td>.17</td>
<td>8,086</td>
<td>.681</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>TerraNova-Reading</td>
<td>-.35</td>
<td>.73</td>
<td>-2.90</td>
<td>8.29</td>
<td>-.03</td>
<td>.12</td>
<td>3,765</td>
<td>.733</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Supera-Reading</td>
<td>-1.61</td>
<td>.11</td>
<td>-10.85</td>
<td>6.76</td>
<td>-.13</td>
<td>2.58</td>
<td>4,319</td>
<td>.112</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>Outcome</td>
<td>t</td>
<td>p</td>
<td>ATT β</td>
<td>SE b</td>
<td>ATT b</td>
<td>F</td>
<td>df</td>
<td>p</td>
<td>Adjusted R²</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>TerraNova-Math</td>
<td>-1.13</td>
<td>.26</td>
<td>-12.75</td>
<td>11.33</td>
<td>-1.13</td>
<td>1.27</td>
<td>2.80</td>
<td>.263</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>TerraNova-Reading</td>
<td>-1.52</td>
<td>.13</td>
<td>-15.46</td>
<td>10.16</td>
<td>-.18</td>
<td>2.32</td>
<td>3.12</td>
<td>.131</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>Supera-Reading</td>
<td>-1.84</td>
<td>.40</td>
<td>-17.43</td>
<td>20.84</td>
<td>-.06</td>
<td>.70</td>
<td>6.91</td>
<td>.402</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>TerraNova-Math</td>
<td>1.36</td>
<td>.18</td>
<td>13.75</td>
<td>10.14</td>
<td>.14</td>
<td>1.84</td>
<td>3.55</td>
<td>.183</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>TerraNova-Reading</td>
<td>-3.88</td>
<td>.71</td>
<td>-3.25</td>
<td>8.69</td>
<td>-0.4</td>
<td>.14</td>
<td>3.77</td>
<td>.714</td>
<td>.00</td>
</tr>
<tr>
<td>2</td>
<td>STAAR-Math</td>
<td>3.33</td>
<td>.00</td>
<td>54.52</td>
<td>26.38</td>
<td>.26</td>
<td>11.07</td>
<td>7.33</td>
<td>.000**</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>STAAR-Reading</td>
<td>3.39</td>
<td>.00</td>
<td>55.27</td>
<td>26.30</td>
<td>.20</td>
<td>11.50</td>
<td>7.33</td>
<td>.000**</td>
<td>.00</td>
</tr>
<tr>
<td>3</td>
<td>STAAR-Math</td>
<td>.92</td>
<td>.36</td>
<td>21.46</td>
<td>23.23</td>
<td>.10</td>
<td>.85</td>
<td>4.35</td>
<td>.364</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>STAAR-Reading</td>
<td>.74</td>
<td>.46</td>
<td>17.66</td>
<td>23.84</td>
<td>.09</td>
<td>.55</td>
<td>4.35</td>
<td>.462</td>
<td>.00</td>
</tr>
<tr>
<td>4</td>
<td>STAAR-Math</td>
<td>2.68</td>
<td>.01</td>
<td>70.72</td>
<td>26.38</td>
<td>.38</td>
<td>7.19</td>
<td>4.92</td>
<td>.014*</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>STAAR-Reading</td>
<td>3.79</td>
<td>.00</td>
<td>100.15</td>
<td>26.42</td>
<td>.40</td>
<td>14.36</td>
<td>4.92</td>
<td>.000**</td>
<td>.00</td>
</tr>
<tr>
<td>5</td>
<td>STAAR-Math</td>
<td>2.11</td>
<td>.04</td>
<td>49.70</td>
<td>23.62</td>
<td>.22</td>
<td>4.43</td>
<td>8.62</td>
<td>.042*</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>STAAR-Reading</td>
<td>1.92</td>
<td>.05</td>
<td>45.21</td>
<td>23.55</td>
<td>.15</td>
<td>3.69</td>
<td>8.62</td>
<td>.039*</td>
<td>.00</td>
</tr>
<tr>
<td>6</td>
<td>STAAR-Math</td>
<td>1.36</td>
<td>.17</td>
<td>30.30</td>
<td>22.24</td>
<td>.13</td>
<td>1.86</td>
<td>8.76</td>
<td>.172</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>STAAR-Reading</td>
<td>1.00</td>
<td>.32</td>
<td>22.53</td>
<td>22.48</td>
<td>.09</td>
<td>1.01</td>
<td>8.76</td>
<td>.323</td>
<td>.00</td>
</tr>
<tr>
<td>7</td>
<td>STAAR-Math</td>
<td>-.96</td>
<td>.34</td>
<td>-38.04</td>
<td>38.35</td>
<td>-0.12</td>
<td>.92</td>
<td>7.55</td>
<td>.341</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>STAAR-Reading</td>
<td>.01</td>
<td>.99</td>
<td>.45</td>
<td>48.38</td>
<td>.00</td>
<td>.00</td>
<td>7.55</td>
<td>.991</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Note. *p less than .05, two tailed. **p less than .001, two tailed. Bold = statistically significant ATT.*

Role of Propensity Score Analysis in the Evaluation

The quality of impact studies is determined based on the research design and methodology. According to MIECHV’s criteria, designation of a high-quality impact study is reserved for random assignment with low attrition rates (Avellar & Supplee, 2013; Avellar et al., 2016). Both criteria are not feasible for a program such as HIPPY, where random assignment design was not applicable as the program enrolls families based on preset criteria, and the attrition rate is high. The current evaluation was designed to overcome the quasi-experimental design limitations to achieve internal and external validity by (a) selecting the pre-treatment...
variables grounded in the literature that play a role in the TX HIPPY program selection and included them in the propensity score estimation model, (b) conducting sensitivity analysis to detect the hidden bias resulting from unmeasured variables, (c) applying the full optimal propensity score matching algorithm that achieves the external validity by using the whole sample frame including all program and non-program participants in the DISD in Grades K to 8, and (d) additionally, the evaluation framework converted the standard scores for each assessment across the grades to z-scores to have them all on the same scale to increase the sample size.

Optimal full matching, despite its limited application in educational studies, has a solid theoretical context and is supported by research findings that prove its effectiveness in reducing the sample selection bias. Optimal full matching is a three-in-one analysis that applies weighing, stratification, and matching; this evaluation highly recommends using optimal full algorithm over other algorithms. Even though optimal full matching is so effective in strengthening the internal and external validity, the complication of the matching analysis, ATT estimation, and its interpretation make other matching algorithms (e.g., one-to-one) more popular. The main strength of this evaluation is the well-formed non-HIPPY group that consists of eligible children for HIPPY participation, in addition to the extremely strong analytical framework to overcome the quasi-experimental design limitations along with the practical diagnostic approach to assess the selection bias in PSM.
<table>
<thead>
<tr>
<th>Grade / Subject</th>
<th>HIPPY</th>
<th>non-HIPPY</th>
<th>MD</th>
<th>95% CI of the difference</th>
<th>t</th>
<th>df</th>
<th>Sig</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Istation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kindergarten</td>
<td>600.52</td>
<td>186.31</td>
<td>616</td>
<td>434.21</td>
<td>252.87</td>
<td>7,472</td>
<td>-166.30</td>
<td>-186.66</td>
</tr>
<tr>
<td>TerraNova</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math - Grade K</td>
<td>788.54</td>
<td>352.05</td>
<td>312</td>
<td>785.16</td>
<td>297.09</td>
<td>5,612</td>
<td>-17.95</td>
<td>-31.95</td>
</tr>
<tr>
<td>Math - Grade 1</td>
<td>796.42</td>
<td>292.13</td>
<td>312</td>
<td>790.53</td>
<td>289.41</td>
<td>5,612</td>
<td>-5.97</td>
<td>-38.98</td>
</tr>
<tr>
<td>Math - Grade 2</td>
<td>765.22</td>
<td>290.10</td>
<td>400</td>
<td>785.52</td>
<td>290.68</td>
<td>6,516</td>
<td>20.30</td>
<td>-9.05</td>
</tr>
<tr>
<td>TerraNova</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading - Grade K</td>
<td>1135.08</td>
<td>95.84</td>
<td>286</td>
<td>1131.75</td>
<td>95.84</td>
<td>3,481</td>
<td>3.67</td>
<td>-8.02</td>
</tr>
<tr>
<td>Reading - Grade 1</td>
<td>1136.12</td>
<td>103.86</td>
<td>158</td>
<td>1137.86</td>
<td>98.91</td>
<td>2,644</td>
<td>-4.47</td>
<td>31.76</td>
</tr>
<tr>
<td>Reading - Grade 2</td>
<td>1139.42</td>
<td>90.06</td>
<td>186</td>
<td>1136.33</td>
<td>95.64</td>
<td>3,351</td>
<td>-3.09</td>
<td>-17.18</td>
</tr>
<tr>
<td>Supera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading - Grade K</td>
<td>1148.56</td>
<td>91.07</td>
<td>330</td>
<td>1163.08</td>
<td>85.17</td>
<td>3,991</td>
<td>14.53</td>
<td>4.91</td>
</tr>
<tr>
<td>Reading - Grade 1</td>
<td>1150.43</td>
<td>86.84</td>
<td>154</td>
<td>1162.55</td>
<td>88.14</td>
<td>2,968</td>
<td>12.12</td>
<td>-2.15</td>
</tr>
<tr>
<td>Reading - Grade 2</td>
<td>1165.24</td>
<td>91.89</td>
<td>214</td>
<td>1165.44</td>
<td>88.48</td>
<td>3,165</td>
<td>.192</td>
<td>-12.09</td>
</tr>
<tr>
<td>STAAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math - Grade 3</td>
<td>1482.15</td>
<td>189.41</td>
<td>344</td>
<td>1428.55</td>
<td>212.53</td>
<td>6,995</td>
<td>-53.60</td>
<td>-76.49</td>
</tr>
<tr>
<td>Math - Grade 4</td>
<td>1570.13</td>
<td>170.02</td>
<td>170</td>
<td>1544.83</td>
<td>215.08</td>
<td>4,183</td>
<td>-25.32</td>
<td>-58.07</td>
</tr>
<tr>
<td>Math - Grade 5</td>
<td>1657.37</td>
<td>172.04</td>
<td>174</td>
<td>1591.47</td>
<td>245.32</td>
<td>4,751</td>
<td>-65.86</td>
<td>-102.65</td>
</tr>
<tr>
<td>Math - Grade 6</td>
<td>1627.15</td>
<td>202.12</td>
<td>192</td>
<td>1578.02</td>
<td>231.12</td>
<td>8,438</td>
<td>-49.13</td>
<td>-82.11</td>
</tr>
<tr>
<td>Grade / Subject</td>
<td>HIPPY Mean</td>
<td>HIPPY SD</td>
<td>HIPPY N</td>
<td>non-HIPPY Mean</td>
<td>non-HIPPY SD</td>
<td>non-HIPPY N</td>
<td>MD</td>
<td>95% CI of the difference</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>----------</td>
<td>---------</td>
<td>----------------</td>
<td>--------------</td>
<td>-------------</td>
<td>----</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Math - Grade 7</td>
<td>1642.75</td>
<td>230.23</td>
<td>218</td>
<td>1609.37</td>
<td>231.31</td>
<td>8,548</td>
<td>-33.42</td>
<td>-64.51</td>
</tr>
<tr>
<td>Math - Grade 8</td>
<td>1487.18</td>
<td>307.72</td>
<td>138</td>
<td>1504.87</td>
<td>316.12</td>
<td>7,414</td>
<td>17.30</td>
<td>-35.98</td>
</tr>
<tr>
<td>Reading - Grade 3</td>
<td>1427.11</td>
<td>199.14</td>
<td>344</td>
<td>1383.12</td>
<td>212.65</td>
<td>6,995</td>
<td>-43.93</td>
<td>-66.89</td>
</tr>
<tr>
<td>Reading - Grade 4</td>
<td>1486.24</td>
<td>194.32</td>
<td>170</td>
<td>1467.14</td>
<td>218.61</td>
<td>4,183</td>
<td>-19.13</td>
<td>-52.52</td>
</tr>
<tr>
<td>Reading - Grade 5</td>
<td>1639.42</td>
<td>189.72</td>
<td>174</td>
<td>1540.68</td>
<td>245.56</td>
<td>4,751</td>
<td>-98.72</td>
<td>-135.62</td>
</tr>
<tr>
<td>Reading - Grade 6</td>
<td>1556.36</td>
<td>210.64</td>
<td>192</td>
<td>1510.65</td>
<td>229.63</td>
<td>8,438</td>
<td>-45.72</td>
<td>-82.11</td>
</tr>
<tr>
<td>Reading - Grade 7</td>
<td>1616.19</td>
<td>220.71</td>
<td>218</td>
<td>1585.62</td>
<td>234.42</td>
<td>8,548</td>
<td>-30.54</td>
<td>-62.01</td>
</tr>
<tr>
<td>Reading - Grade 8</td>
<td>1628.13</td>
<td>228.13</td>
<td>138</td>
<td>1652.58</td>
<td>402.56</td>
<td>7,414</td>
<td>24.45</td>
<td>-42.93</td>
</tr>
</tbody>
</table>

*Note. *p less than .05, two tailed. **p less than .001, two tailed. MD = Mean Difference. Bold = statistically significant ATT.
Evaluation Recommendations

Despite the positive findings, this evaluation suggests that HIPPY children may perform even better if the TX HIPPY management develops a program strategy (long-term and short-term) that focuses on maintaining the program model, the fidelity of the program implementation, and set success measures accordingly. Assessing fidelity is crucial to understanding program impact, and a thorough explanation of how a program was adapted is important in guiding future efforts at successful implementation. This evaluation noted many areas of potential improvements; for instance, the program should consider developing a plan to increase parents’ involvement and interest in the program, developing a plan to enroll the families in two-year program rather than one-year program, and setting a plan to increase parents’ involvement during the summer by providing review packets and supplementary materials to keep the momentum for the next program year.

Additionally, the TX HIPPY home visitors are paraprofessionals typically recruited from the community being served. The paraprofessionals can be more effective than professionals at building rapport with the families because of the reduced social distance (National Academies Press, 1999). Although there is less social distance between the home visitors and the families, paraprofessionals typically have less training or formal education than professionals. The home visitors’ role is critical in maintaining the program model and fidelity of program implementation. An important point to consider is providing ongoing training for the home visitors on how to enroll families, how to administer the standardized assessment, how to score and use the results to meet the child’s needs, how to build rapport with the child before the assessment, how to maintain the parents’ interest in the program, how to conduct the home visits, and how implement the program’s curriculum is a priority. Moreover, the TX HIPPY program
uses the Bracken School Readiness Assessment (BSRA-3) as the measurement tool to monitor children’s educational progress and evaluate program effectiveness; the BSRA-3 involves pre-(at enrollment) and post-assessment (by the end of the program year or when a child leaves the program). Using BSRA-3 pre-test as a diagnostic tool for the home visitors and parents to identify the child’s needs and set an individualized educational plan to meet this need is essential in achieving the program’s pre-set goals.

Furthermore, the TX HIPPY state office made a program adaptation based on previous evaluation finding that the program is more effective for three-year-old children. This adaptation was made without looking at the reasons behind it; the ineffectiveness of the intervention for children who enroll at age four is that HIPPY curriculum three is not standards-based. So, there is a need to consider revising the HIPPY year three instructional materials designed for five-years-old children because the current curriculum is not standards-based and lacks age-appropriate activities. Increasing the knowledge and skills for this age group is the key to improving the effectiveness of the TX HIPPY Program. Finally, in order to maintain program accountability, the home visitors’ role should be delivering the program curriculum for 30 weeks, and the program coordinators’ role is to conduct the pre- and post-assessments. Currently, the home visitors’ role should be to run all program activities related to the home visits (collect the consent form, conduct the pre- and post-assessments, and deliver the HIPPY curriculum). Finally, it is recommended that TX HIPPY make the data and results of the fidelity of implementation available to program developers, implementers, evaluators, and funders to help them improve fidelity assessment and to assess the adoption or change made to the program upon implementation.
Table 9

Descriptive Statistics and Independent Samples t-test Results Comparing HIPPY and non-HIPPY participants on the Z-scores for each Standardized Assessment

<table>
<thead>
<tr>
<th>Grade / Subject</th>
<th>HIPPY</th>
<th>non-HIPPY</th>
<th>Mean Diff.</th>
<th>SE</th>
<th>95% CI of the difference</th>
<th>t</th>
<th>df</th>
<th>Sig</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Istation-School Readiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kindergarten</td>
<td>.64</td>
<td>.75</td>
<td>308</td>
<td>-.03</td>
<td>.26</td>
<td>7,780</td>
<td>.05</td>
<td>-.78</td>
<td>-.55</td>
</tr>
<tr>
<td>TerraNova-Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Grades (K-2)</td>
<td>-.01</td>
<td>1.01</td>
<td>664</td>
<td>.00</td>
<td>.99</td>
<td>18,736</td>
<td>.06</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>TerraNova-Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Grades (K-2)</td>
<td>-.02</td>
<td>1.04</td>
<td>297</td>
<td>.01</td>
<td>1.13</td>
<td>9,189</td>
<td>.06</td>
<td>-.09</td>
<td>.14</td>
</tr>
<tr>
<td>Supera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Grades (K-2)</td>
<td>-.10</td>
<td>1.07</td>
<td>316</td>
<td>.03</td>
<td>1.19</td>
<td>9,658</td>
<td>.06</td>
<td>-.01</td>
<td>.22</td>
</tr>
<tr>
<td>STAAR-Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Grades (3-8)</td>
<td>.096</td>
<td>.83</td>
<td>618</td>
<td>-.00</td>
<td>1.00</td>
<td>40,947</td>
<td>.04</td>
<td>-.18</td>
<td>-.02</td>
</tr>
<tr>
<td>STAAR-Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Grades (3-8)</td>
<td>.06</td>
<td>.72</td>
<td>618</td>
<td>-.00</td>
<td>1.00</td>
<td>40,947</td>
<td>.04</td>
<td>-.14</td>
<td>.02</td>
</tr>
</tbody>
</table>

Note. *p less than .05, two tailed. **p less than .001, two tailed. Bold = statistically significant ATT.
Limitations and Future Research

This evaluation made a choice to use more than one statistical test to examine the treatment effect to overcome some of the limitations. The optimal full and optimal matching was applied to the same sample for each grade level to see which matching algorithm meet the balancing criteria better. Moreover, the attrition rate was high, especially for Grades 3 and up, resulting in a small sample size. The small sample size affects the statistical power to detect the treatment effect and the overlap in the groups’ propensity score distribution. The independent sample t-test on the z-scores for each assessment across the grade levels was conducted to increase the sample size, increase the overlapping between the groups, and to have all the participants on the same scale as the way to overcome this limitation.

However, quite a few more limitations due to methodological and conceptual issues do persist. Assuming the number of covariates (15) is sufficient to check for selection bias is questionable; more than just demographic covariates should be included in the propensity score model (Stuart, 2010). Furthermore, the evaluation did not identify the variables that can be considered as having the joint effect of outcome and treatment among the covariates such as parents’ level of engagement in their child education and the exposure to other educational programs; therefore, the estimation may be biased. The previously mentioned limitations are due to the data availability in the DISD data sources. Additionally, ISIP early reading is an assessment and instruction tool; children who enroll in Pre-K might be exposed to the instruction part, which makes it a possible confounding variable. Participation in other early childhood education programs is also a potential confound. Also, it is worth noting that all statements made about the effectiveness of TX HIPPY program assume fidelity of implementation. However,
there is no fidelity data to test these assumptions. Therefore, all statements about effectiveness should be interpreted with caution.

An interesting extension to this evaluation would be to create a larger dataset of HIPPY and non-HIPPY participants with a complete set of covariates for five consecutive years to see the achievement trends across the years. Future research should continue to track TX HIPPY enrolled families to determine the long-standing effects of participation and to discover additional benefits of the program participation. In contrast, as previous research stated, the gains for intervention programs such as HIPPY may decline over time. Defining the possible loss of program effect would also be a valuable extension.

Furthermore, investigating the levels of HIPPY exposure would make an essential contribution by examining the performance of HIPPY participants who completed one program year and those who completed two program years. Future research should also assess adding more covariates to the propensity score model in addition to defining the variables that have a joint effect of outcome and treatment. Moreover, based on the findings presented in this evaluation, previous optimal full matching studies found meaningful advantages over other methods. Despite the attractions; the application of this method was limited due to the availability of the software. Now that R Programming provides packages for all kinds of propensity scores algorithms, a simulation study is recommended that other methods include propensity score stratification, weighting, and multivariate regression adjustment be applied along with the optimal full propensity score matching to examine the effectiveness of each method in reducing the selection bias and achieve balancing groups.
Conclusion

This evaluation study was intended to estimate the effect of the TX HIPPY intervention program on one of its long-term outcomes, academic achievement in math and reading for Grades K to 8 in addition to the school readiness measure for kindergarten students. HIPPY’s main goal is to empower parents of 3, 4, and 5-year-old children, so they can be involved in their children’s education as primary educators in the home to help them be ready for success in school (HIPPY USA, n.d.). However, its actual impact is continuously questioned and often not rigorously evaluated. Non-randomization, high attrition rates, missing information, and weak evaluation design issues can severely affect the ability to conduct an extensive impact evaluation. This evaluation represents an appropriate solution to tackle missing information and a high attrition rate, and to thoroughly approach the selection bias limitation to make a valid inference.

The evaluation findings show that the TX HIPPY program positively influences kindergarten students to start school ready to learn. HIPPY children were found to present a better performance on math and reading achievement; the difference reached statistical significance on three grade levels (Grades 3, 5, and 6) with a modest positive impact. However, the evaluation findings also indicate that the impact evaluation framework must be designed with attention to the higher-level factors beyond academic achievement that influence children’s academic success.
APPENDIX A

EXTENDED LITERATURE REVIEW
Home Visitation Programs: Impact Evaluation Methodology

The United States is gradually recovering from the Great Recession, the most prolonged phase of economic setback since the Great Depression in the 1930s (Isaacs, 2010). The Great Recession has had a substantial impact on increasing poverty in the United States, and the poverty rate has reached its peak since the early 1980s (DeNavas-Walt et al., 2010). In 2007, more than 13 million children under age 18 were living in poverty (U.S. Census Bureau, 2007), and two years later the number increased by more than two million (DeNavas-Walt et al., 2010). A study conducted in 2009 by Brookings Institute predicted that one million more children would abject poverty after the recession come to an end. More than one in five children lived in poverty in 2009 (Isaacs, 2010). The period between 2007 and 2009 emphasized a major focus on increasing the awareness about family and child poverty challenges, which highlighted the significance of the early childhood years.

During the recession, there were many challenges in the life of pregnant women and children up to the age of five: First, late-preterm infants marked up between 1990 and 2006 by 25%, this brought attention to them as an at-risk population (Martin, Hamilton, Osterman & Shepherd, 2009). Late-preterm babies are susceptible to greater morbidity and mortality rates than full-term infants and are vulnerable to lower cognitive development and less successful school outcomes during childhood (McGowan, Alderdice, Holmes, & Johnston, 2011). Second, based on U.S. Department of Health and Human Services (DHHS) definition, child maltreatment comprises physical and psychological neglect, or sexual abuse that may happen independently or in unison (DHHS, 2010). Research has associated early hardships with increased risks for emerging depression and experiencing poor health later in life (Anthony et al., 2011). Third, family level factors such as unemployment, financial hardship, social isolation, teenage
parenthood, single motherhood, and low educational level have a negative influence on parent-child interaction and children's long-term cognitive, social-emotional, and educational development (Anthony et al., 2011; Evans, 2004).

After the Great Recession, the Patient Protection and Affordable Care Act (PPACA) was signed into law to help guide significant changes, including health care policies, for endangered pregnant women and families. One change was the approval of a one and a half billion evidence-based home visitation initiative to the states over five years, known as the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) Program (Health Resources and Services Administration [HRSA], n.d.). The primary goal of MIECHV is to support families with young children and pregnant women in mastering skills and to obtain resources needed to promote development, health, and school readiness. It is worth noting that MIECHV's insertion in ACA recognized a federal promise to fund evidence-based home visitation programs based on evidence of effectiveness (Adirim & Supplee, 2013). Intervention programs in the field of education are multifaceted. They often include a diagnostic component to identify the needs along with a service component; this complexity increases the challenges for program evaluation (Bowden et al., 2017). Hence, the rigorous impact evaluation of state-level early-childhood intervention programs is a vital priority (Muschkin et al., 2015). The goal of evaluating an intervention program in education is to estimate the treatment effect (Mark & Henry, 2006). The prime research interest in such an evaluation is whether the intervention program improves educational outcomes.

There are several impact evaluation studies that investigate the effectiveness of home visiting programs with underprivileged families that focus on program effectiveness results. There was no step-by-step impact evaluation methodology for home visitation programs that
could be utilized to guide the process of implementing the evaluation research. This article is intended to present a quasi-experimental design methodology by employing a matched comparison group design to create baseline equivalence treatment and comparison groups on selected measures. Notably, it explains the process of propensity score analysis to get an unbiased estimate of the treatment effect to fill the gap in the literature. This work is most important to policymakers, program managers, and evaluators in the field of home visiting interventions in which it guides implementing rigorous evaluation studies.

Evidence-based Home Visitation Programs

Home visiting term covering wide-ranging social programs that use home visits as a method of delivery, implemented by qualified professionals (Barton, 2016; Boller et al., 2010; Sweet & Appelbaum, 2004). The magnitude to which early childhood intervention programs generate long-lasting gains in children's socialization, cognitive development, and school success is a matter of controversy when it comes to the feasibility and cost-effectiveness. Several comprehensive reviews and meta-analyses on the effectiveness of early childhood intervention programs have shown their success in improving children's socioemotional development and parenting knowledge and skills (Beckwith, 2000; Egeland, Weinfield, Bosquet, & Cheng, 2000; Heinicke, Beckwith, & Thompson, 1988; Lojkasek, Cohen, & Muir, 1994; van IJzendoorn, Juffer, & Duyvesteyn, 1995). The 2011-2012 National Survey of Children's Health was employed to assess the national and state frequency of home visits for children up to age 3. The results indicate that about two million children and families received home visits through pregnancy and up to three years, and nationally, about 19% of children received home visits were below the federal poverty level.
Home visitation interventions were used systematically in the United States in the early 19th century (Meckel, 1990), and developed over time through three approaches. First, the public health approach (Boller et al., 2010), which started in the early nineteen century when public health nurses delivered postnatal supplemental visits to check on healthy living conditions in the home. Second, the decrease in child maltreatment, which began in the nineteen-sixties. Third, focusing on debating parent-child relationships and the early childhood education that took shape in the nineteen-seventies. Head Start initiated a home-based program delivery to assist rural needy families; home visitation for families with young children is a long lasting approach proposing risk assessment, guidance, and childcare support interferences at home.

Federal agencies recognized 17 home visiting intervention models in 14 states, including Healthy Families America (HFA), Home Instruction for Parents of Preschool Youngsters (HIPPY), Nurse-Family Partnership (NFP), Early Head Start (EHS), and Parents as Teachers (PAT). The delivery of home visitation is generally meant to enhance children's health and developmental outcomes and parents' capacity and skills (Field, Widmayer, Greenberg, & Stoller, 1982; Krugman, 1993; Logan, 1997). Federally funded intervention programs commonly target low income, limited education families who are at risk for developmental delays, family violence, infant mortality, unequal access to health care, disabilities, and social isolation (Johnson, 2009). In recent years, early childhood home visiting interventions have expanded across the United States, supported by federal policy (Lanier et al., 2015). Some considered the Great Recession as an influential event for leading anti-poverty reform and increasing awareness of child poverty (Colburn, 2014; Isaacs, 2010). The organizational structure of home visiting programs have shifted from locally implemented programs to national models (Boller et al., 2010).
The nationally implemented models underline the devotion of model fidelity and provide a rigorous evaluation of program effectiveness that promotes extraordinary advocacy for strategies related to early childhood through home visiting (Barton, 2010; Boller et al. 2010). Many studies have shown that participation in early childhood education programs has a positive impact on children as they experience gains that persevere throughout their education and into early adulthood (Barnett, 2011; Field et al., 1982; Krugman, 1993; Logan, 1997). Nonetheless, evaluations of large-scale, state-level, and federal programs imply a various pattern regarding the tenacity and cost-effectiveness of their effects (Barnett, 2011; U.S. Department of Health and Human Services, 2010). DHHS supervises the grants and monitors grantees to ensure the fulfillment of the federal requirements for evidence of effectiveness to prove eligibility for funding (HRSA, n.d.; Avellar et al., 2016). The requirements can be summarized as follows: all grantees should recognize at-risk populations where states would spend home visiting funds according to a set of indicators and identify the targeted population of home visitation services (e.g., low-income families, teenage mothers, military families, and child welfare involvement). The states must gather, report, and substantiate progress indicators in eight domains (child health, child development, school readiness, family economic self-sufficiency, linkages and referrals, maternal health, positive parenting practices, and reductions in child maltreatment, juvenile delinquency, family violence, and crime), and at least 75% of the funding must be allocated directly to the program.

A point system has been developed to prioritize the models for review by DHHS and determine the models' eligibility for funding. The points are assigned based on (Avellar & Supplee, 2013; Avellar et al., 2016; HRSA, n.d.): First, the design of impact studies; two points for each matched comparison group design, and three points for each randomized
controlled trial, regression discontinuity, or single case design. Second, the sample size of impact studies; one point for each study with a sample size of 250 or more. Third, the outcome of interest; one point for each study that has an outcome in family economic factors, child maltreatment, juvenile delinquency, and family violence or crime. In a 2012 DHHS review, 32 models were prioritized (including the program models stated previously), of which 12 met the DHHS criteria (Avellar & Supplee, 2013). Similarly, in 2017, DHHS prioritized 45 program models, of which 20 met the DHHS criteria. For the approved evidence-based program models, the process of establishing evidence of effectiveness should be updated every two years (Avellar et al., 2016).

The quality of impact studies is determined based on the research design and methodology of each study. The high-quality impact study is reserved for random assignment with low attrition rates, regression discontinuity, and a single case that meets the What Works Cleaning House (WWC) design criteria. The moderate quality impact study is similarly feasible for randomized assignment studies when they do not attain all criteria for a high rating (e.g., flaws in study design, implementation, and analysis), quasi-experimental design (matched comparison group design that establishes equated groups based on selected measure), regression discontinuity, and single design that meet WWC criteria. Hence, the evidence-based early childhood home visiting program is required to accomplish one of the following: (a) at least one high or moderate quality impact study, finding statistically significant effects in at least two outcome domains; or (b) a high or moderate quality impact study, finding at least one statistically significant effect with non-overlapping samples on the same domain area (Avellar & Supplee, 2013; Avellar et al., 2016). DHHS does not consider low-quality studies as evidence of effectiveness.
Impact Evaluation Methodology for Quasi-experimental Data

Rigorous program evaluation is determined by validly estimating a program's causal impacts (Caliendo & Kopeinig, 2008; Rubin, 1974). The statistical foundation of program evaluation is grounded in the practice of the randomized experiment (Guo & Frasher, 2010). The discussion has been centered on the function of the observational studies and experiments that take part in detecting such impacts. The fundamental justification for favoring randomized experiments is that, when correctly applied, they produce equal treatment and control groups on any measured or unmeasured variables. Mill (1848) described three conditions that must be met in order to claim causal relationships; the cause must precede the effect, a variation in the cause should result in variation in the effect, and the evaluator must be able to eliminate all possible explanations that might affect the variables relationships. Shandish, Cook, and Campbell (2002) stated that to be able to fulfill Mill's causal conditions, we need to implement an experiment that is an empirical investigation to measure the outcome of interest. Randomized experiment satisfies Mill's three conditions for making causal inferences; the random assignment removes all other plausible explanations for any differences in the detected effect. Similarly, the quasi-experiment design is intended to address cause and effect questions (Harder, 2010).

Causal Inference

The purpose of conducting research in education is to see how the outcomes of individuals who receive the treatment differ from those who do not receive the treatment. The counterfactuals are the primary challenge that prevents us from having Mill's condition in the experiments when we do not have a random assignment (Rubin, 1974). Campbell (1957) stated that the threat to internal validity is an important issue to consider when measuring cause and effect. If the threat to internal validity is removed, then we fulfilled Mill's third causal condition
of eliminating the other plausible factors. The randomized experimental design has become the
gold standard in program evaluation introduced by Fisher (1925) to be an effective way of
assessing treatment effects in every field of study. Successful randomization permits other
factors to be ignored since both groups are likely to have the same distribution of the variables,
which produces internal validity. The lack of confounding between the experiment and the
control groups indicates that no factors are confounding the relationship between the treatment
\(W\) and the outcome \(Y\), which is termed as the strong ignorability assumption. However, the
groups might diverge on some characteristic, but the divergence is not significantly associated
with the outcome variable. If this is the case, it would be stated as weak confoundedness or weak
avoidability.

**Quasi-experimental Studies**

In social and behavioral sciences, assigning participants to comparison groups means
deny treatment to those participants, which is in many settings considered unethical, even
illegal. Randomization is not the only way of obtaining useful exogenous variations in treatment
status and consequently identifying the causal impact of a treatment. Sometimes it is possible to
do so with data from a quasi-experiment, or even possible to do so with data from an
observational study. Quasi-experimental designs might give the same results if we have full
awareness and knowledge of the factors that threaten the validity of the study and control them to
get valid conclusions. The terms 'observational study' and 'quasi-experimental design' can be
used interchangeably, which means an empirical investigation that aims at verifying causal
relationships (Shadish, Cook, & Campbell, 2002). Every quasi-experiment may be confounded,
and as a result, has some design weakness. The design may be confounded by mortality, sample
selectivity, maturation, history, instrumentation, testing, or interaction among them. The primary
concern in such designs is often selectivity because if groups differ in the beginning, they may experience different conditions and respond to the treatment differently. In this case, Rubin's Causal Model suggests four approaches to statistically control for confounding factors: adjusting, matching, stratifying, or weighing the treatment and comparison groups so the differences on the other factors would be eliminated or minimized (Holmes, 2014), among which propensity score matching is used most frequently (Austin, 2007; Guo & Frasher, 2010).

Rubin's Causal Model. Rubin causal framework serves as a valuable tool for the evaluation of causal treatment effect. It stresses that individuals selected for treatment or comparison groups have possible outcomes in both conditions. The model assumes that each participant would have two possible outcomes \((Y_{0i}, Y_{1i})\) that represent, correspondingly, to the potential observed outcome in treatment and comparison groups. \(Y_i\) indicates the outcome variable. \(W_i\) is a dichotomous variable, \(W_i = 1\) denote the treatment group participants, and \(W_i = 0\) denote the comparison group participants. The following mathematical model can express the framework:

\[
Y_i = W_i Y_{1i} + (1 - W_i) Y_{0i}
\]

The development of this model has been credited to Neyman (1923) and Rubin (1974, 1978, 1986). The main purpose of this model is that to be able to conclude the causal relationship between the intervention, \(W -i.\) (cause) and the outcome variable (effect) \(Y_i\), the evaluator needs to estimate the outcome of \(Y_i\) under the condition of \(W_i = 0\), and then compare \(Y_{0i}\) with \(Y_{1i}\) to answer the question: To what extent can the observed difference in outcomes between treated and untreated groups be attributed to an intervention, given that all other things are held constant? Considering that the essential problem of causal inference is that the \(Y -0i.\) is not observed (Holland, 1986), Rubin framework holds that the investigator can evaluate the
counterfactual by estimating the difference in the average outcome of the treatment participants $E(Y_1|W = 1)$ and the average outcome of non-participants $E(Y_0|W = 0)$. Hence, the treatment effect ($\tau$) can be calculated as the difference between the means, so long as they are both measurable, and the standard estimator for the average treatment effect can be expressed as follows:

$$\tau = E(Y_1|W = 1) - E(Y_0|W = 0).$$

Causal inference models in science consist of three elements: a scientific theory, a detailed statistical or mathematical model that denotes the theoretical relationships, and empirical data that permit testing and estimating the models (Blalock, 1971; Heckman, 2005). As in every statistical model, it is important to have a good theory underlying the treatment assignment processes (Sobel, 2005), and the covariates selection of the matching and controlling variables should be on theory and literature as well (Rosenbaum, 2005).

The framework provides applied means to evaluate the counterfactual; however, the model has many issues that need to be considered: (a) the model does not take into account the threats to internal validity or covariates, therefore, the evaluator needs to impose additional assumptions to make accountable inferences; (b) the main interest of evaluators in the standard estimator for the average treatment effect is the average outcome of treatment group if they had not participated in the intervention $E(Y_0|W = 1)$, as long as this is an unobservable effect, $E(Y_0|W = 0)$ is used as a substitution for estimation; (c) the classical randomized experiment assumes no selection bias (Cox, 1958; Fisher, 1925), where $E(Y_0|W = 0) = E(Y_1|W = 1)$, and therefore, $(Y_0|W = 1) = E(Y_1|W = 1)$. The average outcome for the treatment group under the condition of the comparison group is not similar to the average outcome of the comparison group. In other words, $(Y_0|W = 1) \neq E(Y_1|W = 1)$ due to selection bias; (d) Rubin extended the
counterfactual framework to make it more general and applicable to observational studies, and the extension was considered a breakthrough (Sobel, 2005). Thus, observational studies comprise complex situations that require a more rigorous approach to data analysis; (e) the standard estimator model evaluates the average treatment effect, overlooking the individual variability.

The principle idea of implementing matching, stratifying, or weighing is to obtain a large number of untreated participants who are comparable to the treated participants in all related pretreatment characteristics (X), so the difference in outcome (Y) between treated and untreated participants is attributed to the intervention program. This paper is interested in presenting matching as the technique to create similar groups at the baseline. Matching can just be applied if the identifying assumptions are invoked based on the detailed and thoughtful understanding of the setup through which selection into treatment group occurs (Blundell, Dearden, & Sianesi, 2005). The first assumption has many terms that developed over time: Unconfoundedness (Rosenbaum & Rubin, 1983), selectthe ion on observables (Barnow, Cain, & Goldberger, 1980), conditional independence, (Lechner, 1999), exogeneity, (Imbens, 2004); and the ignorable treatment effect assignment assumption. The assumption considers the assignment of intervention participants to binary condition (treatment group vs. comparison group) to be independent of the outcome of the comparison group (Y₀) and the outcome of the treatment group (Y₁). Generally, researchers agree that this assumption is essential for program evaluations containing quasi-experimental data. Second, the stable unit treatment value assumption is an a priori assumption presented by Rubin (1980, 1986) as it considers that the value of Y for unit i when exposed to treatment W, is the identical without regard to the process used to assign
treatment \( w \) to unit \( i \), and without regard to the treatment other units receive. This assumption applies to all \( i = 1, \ldots, N \) and all \( w = 1, \ldots, T \).

**Propensity Score Matching**

Propensity score matching (PSM) offers an approach to program evaluation when the randomized trial is not applicable or unethical. PSM is a relatively new technique that has proven useful for estimating and evaluating the causal treatment effect when using non-experimental or observational data (Caliendo & Kopeinig, 2008; Rubin, 1977). PSM has been commonly used to reduce bias in observational and quasi-experimental studies (Joffe & Rosenbaum, 1999; Austin, 2008), and is widely applied to treatment and comparison groups design (Rosenbaum, 2002; Rubin, 2006). There are four models of propensity score analysis for estimating the treatment effect when the randomization is not ethically or practically possible (Becker & Ichino, 2002; Guo & Frasher, 2010; Holmes, 2014). These models estimate the treatment effect on the outcome by controlling for covariates. First, Heckman's sample selection (1979) models the structure of selection, use the conditional probability of exposing to treatment in the estimation of treatment effect, and the use of logistic regression (treatment group, comparison group) that seeks exogenous factors. Second, PSM (Rosenbaum & Rubin, 1983), balances data can be achieved through matching or resampling comparison participants to treatment participants on the probabilities of receiving the treatment, which follows up with bivariate or multivariate analysis as it is the case in the randomized experiments. Fourth, matching estimators (Abadi & Imbens, 2002, 2006) assigns counterfactual for both treatment and comparison participants by using a vector norm with a positive definite matrix. Fifth, propensity score analysis with non-parametric regression (Heckman, Ichimura, & Todd, 1998) compares treatment participants to all comparison participants based on the distance between propensity scores and then non-
parametric regression is applied to estimate the treatment effect on the outcome for the treatment group. PSM is most frequently used among the four models (Austin, 2007); the motivation for PSM is to shift from ATT to ATE, to eliminate the underlying discrepancies between treatment and comparison groups and remove bias that caused the groups to undergo different treatments.

PSM procedures. The process of propensity score analysis (Guo & Frasher, 2010) starts with propensity score estimation, choosing a matching algorithm, checking for data balancing, assessing the quality of effect estimation and their standard errors, and conducting the sensitivity analysis.

Propensity Score Estimation. The propensity score is defined as the conditional probability that a participant will be in the treatment group, given his or her characteristics (Guo & Frasher, 2010; Holmes, 2014). Propensity scores allow for controlling most, if not all, pretreatment variables, which increases the internal validity of the evaluation design so that the estimation of the treatment effect will be unbiased and accurate (Holmes, 2014; Rubin, 1974). The predicted probability of being in a treatment or comparison group for each participant in the sample is the result of the analysis (Rosenbaum & Rubin, 1984).

In quasi-experimental data, propensity scores can function in different ways. They might be used to adjust data or create matched samples and weights for balancing characteristics between groups or used as control variables (Holmes, 2014). There are four statistical methods with which propensity scores are frequently estimated: logistic regression (Rosenbaum & Rubin, 1983), regression with a dummy dependent variable, discriminant function analysis, and probit regression. In general, there are no criteria for favoring one method over another, but it is recommended that the evaluator use the same approach for estimating the propensity scores and treatment effect (Holmes, 2014). All variables that are measured before the treatment should be
incorporated in the initial propensity scores model to get an unbiased and consistent estimate of
treatment effect regardless of the statistical method used (Guo & Frasher, 2010). In the case of
matched comparison group design, propensity scores are commonly estimated using logistic
regression. Logistic regression permits, based on the characteristics calculation for each
participant, the propensity score that the participant received during treatment (Jupiter, 2017).

Matching as Controlling. Matching is one way of removing pretreatment differences, so
the groups begin to have the same characteristics (Rubin, 1974). Heckman and Navarro-Lozano
(2004) identified the principal issue of matching as selecting the right variables to use because
leaving out significant confounding factors would result in biased and insufficient analysis
procedures. Rubin's (1974) simulations illustrated that matching works better to estimate true
causal effects than regression adjustment. Matching aims at forming similar treatment and
comparison groups by pairing participants who have close to identical propensity scores
(Rosenbaum, 2002; Rubin, 2006).

Propensity score matching is an efficient way of matching, as it resolves the
dimensionality issue when many variables are used to match participants (Holmes, 2014). It
decreases the number of matching variables to a single dimension (Rosenbaum & Rubin, 1983).
The program evaluator needs to consider a set of factors for selecting confounders (Holmes,
2014). The evaluator also needs to choose theories supported by prior research for more
manageable procedures and consider possible untested confounding factors as well, and
empirically test them to eliminate the uncontrolled theoretical confounder possibility.
Additionally, pretest measures of the outcome are commonly used as a separate control variable.
It is theoretically straightforward to think of each participant in one group associated with a
single participant in another group; many statistical procedures are easier to implement when
there is the same number of cases to be in each group. One-to-one matching is statistically adequate and sufficient if the distribution of the propensity scores is symmetrical and unimodal. Nonsymmetrical distributions of the propensity scores mean adding more information (match more participants) might result in sufficient statistical procedures to estimate the causal effect. Greedy one-to-k matching is preferred over one-to-one matching when one has a moderate number of treatment participants. A treatment participant is matched with as many comparison participants as possible, and all of them are removed as probable matches for other participants. Subsequently, treatment participants can only be matched with the leftover, which increases the precision of the effect estimation in the comparison group as it decreases the standard error of the coefficient (percentage, mean, ratio, or rate).

Conversely, it can increase the error variance of the coefficient in the treatment group and decrease the number of matched treatment participants if they are dropped because an earlier participant took their match. Additionally, it is less effective in reducing the imbalance between treatment and comparison groups, as the additional participants matched are not exact matches. Furthermore, the calculation process is time-consuming.

Nearest neighbors matching follows the same method of one-to-one matching except that, instead of picking only the one nearest neighbor, the statistical program selects k nearest neighbors (Rosenbaum, 1989; Rubin, 1979). For nearest neighbors matching to be statistically adequate, there should be only one nearest neighbor propensity, and that neighbor should not be nearest to another treatment propensity. The one-to-k or caliper matching limits how different the participants are allowed to be, as it defines a caliper and selects all matches within the caliper (Austin, 2008; Gu & Rosenbaum, 1993). Caliper matching is statistically sufficient when the caliper is close to exact matching that no closer matching would enhance the balance between
the groups. Kernel matching is a process of matching participants with an identical propensity score and averaging the effect estimate for those participants (Becker & Ichino, 2002).

In optimal or many-to-many matching, each treatment or comparison participant is matched with as many matches as possible within a specified criterion. It is a sequential procedure to select multiple treatment participants for each comparison participant and multiple comparisons for each treatment participant if there are multiple participants that meet the matching criterion (Guo & Frasher, 2010; Holmes, 2014). The theory of optimal full matching assumes that all the untreated participants are used and that there are no restrictions on the number of untreated participants that may be matched with a treated participant (Holmes, 2014; Stuart & Green, 2008). This approach estimates the treatment effect with high precision because it increases the number of untreated participants matched to each treated participant. The increase in the sample size will probably enhance the accuracy of the treatment effect, which mirrors the traditional variance bias trade-off (Guo & Fraser, 2010).

Nevertheless, this strategy might increase bias in estimating treatment effects due to matching dissimilar participants (Shadish & Steiner, 2010). Optimal full matching is a sufficient and statistically adequate procedure; it uses all the available information to minimize the dissimilarity, imbalance, and total distance between matched cases (Holmes, 2015; Guo & Frasher, 2010; Rosenbaum, 1989; Rubin, 1979). It constantly has an optimal solution; no better solution can be found for the available information (Guo & Frasher, 2010; Holmes, 2014; Jupiter, 2017). Optimizing the propensity score minimizes the overall differences, but some participants' covariates may remain imbalanced. Optimal full matching may be implemented using propensity scores alone, with distance measures (e.g., Mahalanobis distance) or with both
propensities and distance measures (Guo & Frasher, 2010; Holmes, 2014; Rosenbaum & Rubin, 1983; Sekhon, 2011).

Data Balancing. After matching has been achieved, the check should always be conducted to make sure that the balance is enhanced between the groups in terms of the observed variables that are used for matching (Jupiter, 2017). Data balancing to control for covariates is a must for conducting meaningful evaluations; it answers an important question about whether the matching technique was able to balance the distribution of related covariates (Caliendo & Kopeinig, 2008). Data are balancing displays the difference between the treatment and comparison groups on the outcome variable if the evaluator enforces them to follow the same covariates (e.g., age, sex, SES) distribution (Guo & Frasher, 2010). After balancing, the covariate is no longer a confounding variable. Data balancing is achieved when a propensity score produces matched groups that have similar distributions on the propensity scores and the covariate measures (Rosenbaum & Rubin, 1984; Shadish, Luellen, & Clark, 2006). Rosenbaum and Rubin (1983) suggested the use of balancing scores $b(X)$, which is the probability of participating in the intervention program, given observed characteristics ($X$). PSM is balancing measures in a way that it can be used to balance the characteristics distribution of two groups so that they are the same (Rosenbaum & Rubin, 1985), and offers a consistent, uniform, and standardized process for matching participants (Guo & Frasher, 2010). PSM reduces bias across participants, known as equal percent bias reduction (Rosenbaum & Rubin, 1985; Rubin, 1976).

Standardized mean difference statistic ($d$) is calculated to examine the imbalance in covariate variables and then, subsequently, for examining the balance achieved by the propensity score technique selected for assessing the treatment effect (Shadish & Steiner, 2010). An index value closest to zero is an indication of balancing to remove residual bias (Ho et al., 2007).
Additionally, Rubin (2001) proposes a set of other indicators: the variance of the propensity score should be close to 1 in the two groups, and the variances ratio of the covariates after adjusting the groups for the propensity score must be close to 1. The rule of thumb to evaluate data balancing is to have the variance ration of 2, with values less than .5 and greater than 2.0 considered extreme (Rubin, 2001). Moreover, many evaluators use $d$ less than .1 or .25 as an indication of data balancing.

Sometimes evaluators cannot attain the necessary degree of balance on all observed variables because of the absence of overlap between treatment and comparison groups, signifying that some participants of the treatment group display characteristics not found in the comparison group and vice versa (Caliendo & Kopeinig, 2008; Holmes, 2014). The two distributions of the treatment and comparison groups should overlap (common support area) greatly for the propensity score analysis to be successful (Caliendo & Kopeinig, 2008). If matching results were not successful, the evaluator needs to go back to the first step and consider different covariates to be incorporated in the model. The concept of well-behaved data is crucial for understanding when greedy matching is approximately as good as optimized matching.

There are many statistical packages like SPSS and SAS, that can be used to implement the procedures of propensity score analysis. For the balancing tests, different macros for matching are available for greedy matching in SPSS, greedy matching in SAS (Parsons, 2001), and optimal matching in SAS (Kosanke & Bergstrahl, 2004). Additionally, Stata statistical software offers packages for propensity score analysis, like PSMATCH2 (Leuven & Sianesi, 2018), MATCH (Abadie, Drukker, Herr, & Imbens, 2004), or PSORE (Becker & Ichino, 2002). R software provides specialized packages as well: MatchIt (Ho et al., 2011), Matching (Sekhon,
Effect Estimation and Their Standard Errors. The treatment effect has two categories depending on the quantity being estimated: Average Treatment Effect (ATE) and Treatment Effect on the Treated (ATT) (Caliendo & Kopeinig, 2008; Holmes, 2014). The mean difference between the treatment and comparison groups in randomized experiment estimates the ATE, while the effect only on those who received the treatment estimates the ATT. In the randomized experiment, ATE and ATT are identical, whereas they differ in nonrandomized experiments. The methods presented in this paper can estimate both; however, it is important that the evaluator knows which treatment effect is being estimated to consider proper procedures for each estimate (Shadish & Steiner, 2010). There are four methods of estimating treatment effect: Ordinary Least Square (OLS) Regression, matching, stratification, and weighing. In the case of a matched-pairs design, a mechanism for propensity score matching is usually used to balance the data by matching each treated participant \( (x_i | w_i = 1) \) to non-treated participant \( (x_i | w_i = 0) \) on the matching variable \( (x) \), and then, comparing the average of the outcome variable \( (Y) \) for the treated and untreated participants. The difference is an estimate of the average treatment effect \( (\tau) \).

\[
\hat{\tau}_{match} = E(\hat{y}_{match,1}|W_{match} = 1) - E(\hat{y}_{match,0}|W_{match} = 0)
\]

Where:

- \( W_{match} = 1 \) is the treatment group with matched participants.
- \( W_{match} = 0 \) is the matched comparison group.

Sensitivity Analysis to Detect Selection Bias. The last step in PSM analysis is the sensitivity assessment of the results concerning the deviations from the pre-identified assumption. Bias
estimation can be extensive if the evaluator neglects important covariates in data from quasi-
experimental studies. Data balancing is the central issue in quasi-experiment data. Even though
evaluators control for selection bias through matching, they can only adjust for the observed
covariates (Blalock, 1971; Ho et al., 2007; Imai, King & Stuart, 2008; Imai & Van Dyk, 2004;
Rosenbaum, 2009; Rubin & Thomas, 2000). Hence, selection bias due to unmeasured covariates
remains a problem (Guo & Frasher, 2010; Rosenbaum, 2002, 2005). Hidden and overt biases that
affect causal inference are well controlled in the randomized experiment, making the error term
in the regression model equal to zero (Guo & Frasher, 2010). Overt bias can arise when there is a
disproportion among the observed covariates, and hidden bias could occur when unobserved
variables are confounders that result in violating the ignorability assumption (Rosenbaum, 2002,
2009). Rubin recommended carrying out sensitivity analysis and examining different sets of
covariates to address the limitation of an inability to adjust for unobserved covariates (Guo &
Frasher, 2010).

Rosenbaum (1991) developed the framework that computes the hidden covariates bias. The
framework defines two odds: The odds ($\Gamma$) for a treated participant and the odds ($\Gamma$) for a
comparison participant. Sensitivity analysis starts with selecting a sequence of odd ratio ($\Gamma$)
values starting at 1, then conducting the McNemar's test (for binary outcomes) or the Wilcoxon
signed rank test (for continuous outcomes) for each ratio ($\Gamma$). Finally, the upper bound of the p-
value of the test statistic is examined for changes over the series of $\Gamma$ values. If the $\Gamma$ value
changed from significant to non-significant, the treatment effect is considered sensitive. If no
change has occurred, the estimated effect is very robust to hidden bias.
Discussion

Policy makers have stressed the importance of systematic development and evaluation of early childhood intervention programs that aim at motivating the young generation to their developmental and educational capabilities and improving their life expectations. Intervention programs in the field of education are complex. Therefore, rigorous impact evaluation of state-level early childhood intervention programs is a critical need. The goal of evaluating an intervention program in education is to estimate the treatment effect by answering the question of whether or not the intervention improves educational outcomes. There are many research designs and statistical tools available to achieve this purpose, but which one yields precise and unbiased treatment effect estimation? While many systematic reviews, evaluations, and meta-analyses exist evaluating the effectiveness of home visiting programs, no previous literature has explained the applied methodology. This article reflected on the process of conducting the evaluation research for quasi-experiment matched pairs design.

For more than three decades, the PSM method has been available to support the testing of cause and effect hypotheses in quasi-experimental data. Nevertheless, PSM techniques up to now have not been applied extensively in the evaluation. This may be due to a lack of knowledge of the fundamental principles and practical guidelines of implementation, or/and due to the availability of the software that carries out the analysis. PSM analysis provides evaluators with an incorporated framework where the overt bias in the estimate of a treatment effect from measured covariates is adjusted, and hidden bias from unobserved covariates can be evaluated concerning the sensitivity of the effect estimate. PSM has been widely used to reduce bias in observational and quasi-experimental studies and is widely applied to treatment and comparison groups design. As yet, there are, in combination, about 15 propensity score estimation and
application techniques to choose from to conduct a matched comparison group evaluation in which the PSM is commonly used to evaluate the causal impact of a treatment effect. The focus of the whole process is to create a sample in which the selection of treatment is not confounded with measured covariates at the baseline to estimate the treatment effect.
APPENDIX B

R-CODE FOR PROPENSITY SCORES ANALYSIS
#Install Packages
install.packages("survival")
library(survival)
install.packages("MatchIt")
library(MatchIt)
install.packages("optmatch")
library(optmatch)
install.packages("Zelig")
library(Zelig)
install.packages("Matching")
library(Matching)
install.packages("cobalt")
library(cobalt)
install.packages("rbounds")
library(rbounds)
install.packages("MKmisc")
library(MKmisc)

## KINDERGARTEN TERRANOVA MATH
imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Math00.csv")
View(imp.data)

## Full Propensity scores Estimation
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file =
    "C:/Users/>>>>/Desktop/Terrafullmath00.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continuous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.thresho = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, get0ption("digits" - 3)))
## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status",
"Primary_disability", "dyslexia", "atrisk", "homeless", "grade"),
"Grade"))
##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp =
".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs =
TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))
# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100")
## MATH ATT estimation

```
z.att0.mi <- zelig(MATH_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att0.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att0.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, TerraNova MATH, Grade K")
combine_coef_se(z.att0.mi)
```

# Sensitivity Analysis

```
Y <- imp.data$MATH_SCORE
tr <- imp.data$HIPPY
psens(Y[tr == "0"], Y[tr == "1"], Gamma=1.7, GammaInc=0.5)
hlsens(Y[tr == "0"], Y[tr == "1"], pr = .1, Gamma=1.7, GammaInc=0.5)
```

## Optimal Propensity scores Estimation and Matching

```
imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Math00.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
  student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
  lunch_status + Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,],
  family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
  country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
  Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], method =
  "optimal", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)
write.csv(imp.data, file =
  "C:/Users/>>>>/Desktop/Terraoptimalmath00.csv")
imp.data[1:10,]
bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".".imp",
  imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
  s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
  TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
  "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, get0ption("digits" - 3)))
```

## visual inspection:
## Using alternate variable names

## Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Balance Plot of Propensity Score Matching:
bal.plot(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", which.imp = 100, imp.summary = FALSE, var.name = "race")

#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)

## list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100")

## Math ATT estimation for the 100 imputed datasets
y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight")

# Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctrl1 <- imp.data$match.weight == "0"
psens(ctrl1, tr1, Gamma=1.7, GammaInc=0.5)
## GRADE 1 TERRANOVa MATH

imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Math01.csv")
View(imp.data)

## Full Propensity scores Estimation and Matching

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
  student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
  lunch_status + Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,],
  family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
  country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
  Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], method = "full",
  ratio = 1, replace = TRUE)

  imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file =
  "C:/Users/>>>>/Desktop/Terrafullmath01.csv")
imp.data[1:10,]

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching",
imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, get0ption("digits" - 3)))

## visual inspection:

## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status",
"Primary_disability", "dyslexia", "atrisk", "homeless", "grade"),
## Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100")

## MATH ATT estimation for the 100 imputed datasets
z.att0.mi <- zelig(MATH_SCORE~ HIPPY, data = mi.out, model = "ls")
ATT(z.att0.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att0.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, TerraNova MATH, Grade 1")
combine_coef_se(z.att0.mi)

# Sensitivity Analysis
Y <- imp.data$MATH_SCORE
tr <- imp.data$HIPPY
psens(Y[tr == "0"], Y[tr == "1"], Gamma=1.7, GammaInc=0.5)
hlsens(Y[tr == "0"], Y[tr == "1"], pr=.1, Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {
  in.imp <- imp.data$imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,] 
family="binomial"$fitted.values 
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang 
+ country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + 
Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = 
"optimal", ratio = 1, replace = TRUE) 

imp.data$match.weight[in.imp] <- m.out$weights 
}

summary(m.out)
write.csv(imp.data, file = 
"C:\Users/>>>>/Desktop/Terraoptimalmath01.csv")
imp.data[1:10,]
bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", 
imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE, 
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = 
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = 
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection: 
## Using alternate variable names 

v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang 
+ country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status", 
"Primary_disability", "dyslexia", "atrisk", "homeless", "grade"), 
new = c("Race", "Ethnicity", "Sex", "Limited English Proficiency", "Primary 
Language", "Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low 
SES Status", "Lunch Status", "Disability Stat", "Dyslexia", "At-risk Student", "Homelessness", 
"Grade") 

##Love Plot of Propensity Score Matching: 
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = 
".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = 
TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Define the multiple imputed data 
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data) 

##list of multiple imputed datasets. 
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, 
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, 
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, 
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, 
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, 
104
m.data$'35', m.data$'36', m.data$'37', m.data$'38', m.data$'39', m.data$'40', m.data$'41', m.data$'42', m.data$'43', m.data$'44', m.data$'45', m.data$'46', m.data$'47', m.data$'48', m.data$'49', m.data$'50', m.data$'51', m.data$'52', m.data$'53', m.data$'54', m.data$'55', m.data$'56', m.data$'57', m.data$'58', m.data$'59', m.data$'60', m.data$'61', m.data$'62', m.data$'63', m.data$'64', m.data$'65', m.data$'66', m.data$'67', m.data$'68', m.data$'69', m.data$'70', m.data$'71', m.data$'72', m.data$'73', m.data$'74', m.data$'75', m.data$'76', m.data$'77', m.data$'78', m.data$'79', m.data$'80', m.data$'81', m.data$'82', m.data$'83', m.data$'84', m.data$'85', m.data$'86', m.data$'87', m.data$'88', m.data$'89', m.data$'90', m.data$'91', m.data$'92', m.data$'93', m.data$'94', m.data$'95', m.data$'96', m.data$'97', m.data$'98', m.data$'99', m.data$'100')

## Math ATT estimation for the 100 imputed datasets

y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight")
# Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

###GRADE 2 TERRANOVA MATH
## Full Propensity scores Estimation and Matching

imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Math02.csv")
View(imp.data)

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)

  imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/Terrafullmath02.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status",
"Primary_disability", "dyslexia", "atrisk", "homeless", "grade"),
new = c("Race", "Ethnicity", "Sex", "Limited English Proficiency", "Primary
Language", "Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low
SES Status", "Lunch Status", "Disability Stat", "Dyslexia", "At-risk Student", "Homelessness",
"Grade"))

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp
= ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs =
TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Balance Plot of Propensity Score Matching:
bal.plot(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp", which.imp = 100, imp.summary = FALSE, var.name = "race")

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
106
m.data$'77', m.data$'78', m.data$'79', m.data$'80', m.data$'81', m.data$'82', m.data$'83', m.data$'84', m.data$'85', m.data$'86', m.data$'87', m.data$'88', m.data$'89', m.data$'90', m.data$'91', m.data$'92', m.data$'93', m.data$'94', m.data$'95', m.data$'96', m.data$'97', m.data$'98', m.data$'99', m.data$'100')

## MATH ATT estimation for the 100 imputed datasets
z.att2.mi <- zelig(MATH_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att2.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att2.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, TerraNova MATH, Grade 2")
combine_coef_se(z.att2.mi)

# Sensitivity Analysis
Y <- imp.data$MATH_SCORE
tr <- imp.data$HIPPY
psens(Y[tr == "0"], Y[tr == "1"], Gamma=1.7, GammaInc=0.5)
hlsens(Y[tr == "0"], Y[tr == "1"], pr= .1, Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Math02.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {
  in.imp <- imp.data$imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file =
  "C:/Users/>>>>/Desktop/Terraoptimalmath02.csv")
imp.data[1:10,]
bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

### visual inspection:
### Using alternate variable names

## Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)

## list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100")

## Math ATT estimation for the 100 imputed datasets
y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight")

# Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctrl1 <- imp.data$match.weight == "0"
psens(ctrl1, tr1, Gamma=1.7, GammaInc=0.5)

108
imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Read00.csv")
View(imp.data)

## Full Propensity scores Estimation and Matching
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
in.imp <- imp.data$.imp == i
imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,],
family="binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
atrisk + homeless , data = imp.data[in.imp,], method = "full", ratio = 1,
replace = TRUE)

imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)
write.csv(imp.data, file =
"C:/Users/>>>>/Desktop/Terrafullread01.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp =
."imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "Primary_disability", 
"dyslexia", "atrisk", "homeless", "grade"),
"Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low SES Status",
"Disability Status", "Dyslexia", "At-risk Student", "Homelessness", "Grade"))
##Love Plot of Propensity Score Matching:
## Reading ATT estimation for the 100 imputed datasets

Y1 <- (imp.data$READ_SCORE)

z.att1.mi <- zelig(READ_SCORE ~ HIPPY, data = mi.out, model = "ls")

ATT(z.att1.mi, treatment = "HIPPY")

qi0 <- get_qi(z.att1.mi, qi = "ATT", xvalue = "TE")

hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, TerraNova READING, Grade 1")

## Sensitivity Analysis

Y <- imp.data$READ_SCORE

tr <- imp.data$HIPPY

psens(Y[tr == "0"], Y[tr == "1"], Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching

imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Read00.csv")

View(imp.data)

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))

for (i in unique(imp.data$imp)) {
  in.imp <- imp.data$imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,]
family="binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)

imp.data$match.weight[in.imp] <- m.out$weights

summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/Terraoptimalread01.csv")

imp.data[1:10,]

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)

print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))

View(m.data)

#list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
## Reading ATT estimation for the 100 imputed datasets

```r
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")
```

### Sensitivity Test

```r
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)
```

### GRADE 1 TERRANOVA READING

#### Full Propensity scores Estimation and Matching

```r
imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Read01.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
                             student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
                             Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,],
                             family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
                   country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
                   atrisk + homeless, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}
```
data = imp.data, weights = "match.weight", method = "matching", imp = ".imp",
imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continuous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)

print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "Primary_disability",
"dyslexia", "atrisk", "homeless", "grade"),

"Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low SES Status",
"Disability Stat", "Dyslexia", "At-risk Student", "Homelessness", "Grade"))

## LovePlot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))

View(m.data)

mi.out <- to_zelig_mi(m.data$'1', m.data$'2', m.data$'3', m.data$'4', m.data$'5', m.data$'6',
m.data$'7', m.data$'8', m.data$'9', m.data$'10', m.data$'11', m.data$'12', m.data$'13',
m.data$'14', m.data$'15', m.data$'16', m.data$'17', m.data$'18', m.data$'19', m.data$'20',
m.data$'21', m.data$'22', m.data$'23', m.data$'24', m.data$'25', m.data$'26', m.data$'27',
m.data$'28', m.data$'29', m.data$'30', m.data$'31', m.data$'32', m.data$'33', m.data$'34',
m.data$'35', m.data$'36', m.data$'37', m.data$'38', m.data$'39', m.data$'40', m.data$'41',
m.data$'42', m.data$'43', m.data$'44', m.data$'45', m.data$'46', m.data$'47', m.data$'48',
m.data$'49', m.data$'50', m.data$'51', m.data$'52', m.data$'53', m.data$'54', m.data$'55',
m.data$'56', m.data$'57', m.data$'58', m.data$'59', m.data$'60', m.data$'61', m.data$'62',
m.data$'63', m.data$'64', m.data$'65', m.data$'66', m.data$'67', m.data$'68', m.data$'69',
m.data$'70', m.data$'71', m.data$'72', m.data$'73', m.data$'74', m.data$'75', m.data$'76',
m.data$'77', m.data$'78', m.data$'79', m.data$'80', m.data$'81', m.data$'82', m.data$'83',
m.data$'84', m.data$'85', m.data$'86', m.data$'87', m.data$'88', m.data$'89', m.data$'90',
m.data$'91', m.data$'92', m.data$'93', m.data$'94', m.data$'95', m.data$'96', m.data$'97',
m.data$'98', m.data$'99', m.data$'100')

## Reading ATT estimation for the 100 imputed datasets
Y1 <- (imp.data$READ_SCORE)

z.att1.mi <- zelig(READ_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att1.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att1.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, TerraNova READING, Grade 1")

# Sensitivity Analysis
Y <- imp.data$READ_SCORE
tr <- imp.data$HIPPY
psens(Y[tr == "0"], Y[tr == "1"], Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching

imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Read00.csv")
View(imp.data)

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
                             student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
                             Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,],
                             family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
                   country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
                   atrisk + homeless , data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)

  imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file =
  "C:/Users/>>>>/Desktop/Terraoptimalread01.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
         atrisk + homeless ,
    data = imp.data, weights = "match.weight", method = "matching", imp = ".imp",
  imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names

114
##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching",
    imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1,
    stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE,
    var.names = v, no.missing = TRUE, limits = c(0, 1))
#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
##list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100`)
## Reading ATT estimation for the 100 imputed datasets
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")
## Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)
### GRADE 2 TERRANOVA READING
### Full Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Read02.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
in.imp <- imp.data$imp == i
imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,],
family="binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
atrisk + homeless, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file =
"C:/Users/>>>>/Desktop/Terrafullread02.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp =".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continuous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))
## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "Primary_disability",
"dyslexia", "atrisk", "homeless", "grade"),
"Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low SES Status",
"Disability Stat", "Dyslexia", "At-risk Student", "Homelessness", "Grade"))
##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching",
imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1,
stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE,
var.names = v, no.missing = TRUE, limits = c(0, 1))

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$imp))
View(m.data)
mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100`)

## Reading ATT estimation for the 100 imputed datasets
Y1 <- (imp.data$READ_SCORE)
z.att2.mi <- zelig(READ_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att2.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att2.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Averege Treatment Effect of HIPPY on the Treated, Optimal Full Matching, TerraNova READING, Grade 2")

# Sensitivity Analysis
Y <- imp.data$READ_SCORE
tr <- imp.data$HIPPY
psens(Y[tr =="0"], Y[tr =="1"], Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/TerraNova.Read02.csv")
View(imp.data)

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {
  in.imp <- imp.data$imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
summary(m.out)
write.csv(imp.data, file =
        "C:/Users/>>>>/Desktop/Terraoptimalread02.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
attrisk + homeless,
        data = imp.data, weights = "match.weight", method = "matching", imp = ".imp",
        imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
        TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.ratio = "as.is", digits = max(3, get0ption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
        "country_of_birth", "spedflag", "TAG_participant", "Low_SES", "Primary_disability",
        "dyslexia", "attrisk", "homeless", "grade"),
        "Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low SES Status",
        "Disability Stat", "Dyslexia", "At-risk Student", "Homelessness", "Grade"))
##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp =
        ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diff.s", abs =
        TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
##list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
## Reading ATT estimation for the 100 imputed datasets

```r
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")
```

# Sensitivity Test
```r
tr1 <- imp.data$match.weight == "1"
ctrl1 <- imp.data$match.weight == "0"
psens(ctrl1, tr1, Gamma=1.7, GammaInc=0.5)
```

### KINDERGARTEN SUPERA READING

## Full Propensity Scores Estimation and Matching
```r
imp.data <- read.csv("C:/Users/>>>>/Desktop/supera00.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
                               student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
                               lunch_status + dyslexia + atrisk + homeless, data = imp.data[in.imp,],
                               family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
                   country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia +
                   atrisk + homeless, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}
```

```r
summary(m.out)
write.csv(imp.data, file =
           "C:/Users/>>>>/Desktop/supera00full.csv")
imp.data[1:10,]
```

```r
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
        country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia +
        atrisk + homeless,
        data = imp.data, weights = "match.weight", method = "matching", imp = ".imp",
        imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continuous = "std", disp.v.ratio = TRUE, un = TRUE,
        s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
        TRUE, disp.means = TRUE)
```
## Reading ATT estimation for the 100 imputed datasets

Y <- imp.data$READ_SCORE
z.att0.mi <- zelig(READ_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att0.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att0.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, Supera READING, Grade K")
combine_coef_se(z.att0.mi)

# Sensitivity Analysis
Y <- imp.data$READ_SCORE
tr <- imp.data$HIPPY
psens(Y[tr =="0"], Y[tr =="1"], Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/supera00.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
  student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
  lunch_status + dyslexia + atrisk + homeless,
  data = imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
  country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia +
  atrisk + homeless,
  data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)
write.csv(imp.data, file =
  "C:/Users/>>>>/Desktop/supera00optimal.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia +
atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp",
imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status", "dyslexia",
"atrisk", "homeless", "grade"),
  "Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low
  SES Status", "Lunch Status", "Dyslexia", "At-risk Student", "Homelessness", "Grade"))

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)

##list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100`)

## Reading ATT estimation for the 100 imputed datasets
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")

# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

### GRADE 1 SUPERA READING
### Full Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/supera.01.csv")
View(imp.data)

imp.data$sps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$sps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
   student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
   lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp],
   family="binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)

imp.data$match.weight[in.imp] <- m.out$weights

}

summary(m.out)
write.csv(imp.data, file =
"C:/Users/>>>>/Desktop/supera01full.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continuous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
"Primary_disability", "dyslexia", "atrisk", "homeless", "grade"),
"Grade"))

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,

123
## Reading ATT estimation for the 100 imputed datasets

\[
Y \leftarrow (\text{imp.data$READ\_SCORE})
\]

\[
z.\text{att1.mi} \leftarrow \text{zelig(READ\_SCORE \sim HIPPY, data = mi.out, model = "ls")}
\]

\[
\text{ATT(z.\text{att1.mi, treatment = "HIPPY")}
\]

\[
\text{Qi1} \leftarrow \text{get\_qi(z.\text{att1.mi, qi = "ATT", xvalue = "TE")}
\]

\[
\text{hist(qi1, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, Supera READING, Grade 1")}
\]

\[
\text{combine\_coef\_se(z.\text{att1.mi)}
\]

# Sensitivity Analysis

\[
Y \leftarrow -\text{imp.data$READ\_SCORE}
\]

\[
\text{tr} \leftarrow -\text{imp.data$HIPPY}
\]

\[
\text{psens(Y[tr =="0"], Y[tr =="1"], Gamma=1.7, GammaInc=0.5)}
\]

## Optimal Propensity scores Estimation and Matching

\[
\text{imp.data} \leftarrow \text{read.csv ("C:/Users/>>>>/Desktop/supera.01.csv")}
\]

\[
\text{View(imp.data)}
\]

\[
\text{imp.data$ps} \leftarrow -\text{imp.data$match.weight} \leftarrow \text{rep(0, nrow(imp.data))}
\]

\[
\text{for (i in unique(imp.data$.imp))}
\]

\[
\text{in.imp} \leftarrow -\text{imp.data$.imp = i}
\]

\[
\text{imp.data$ps[in.imp] \leftarrow glm(HIPPY \sim race + fedEthnicity + Sex + LEP\_program + student\_primary\_lang + country\_of\_birth + spedflag + TAG\_participant + Low\_SES + lunch\_status + Primary\_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], family="binomial")$fitted.values}
\]

\[
\text{m.out} \leftarrow \text{matchit(HIPPY \sim race + fedEthnicity + Sex + LEP\_program + student\_primary\_lang + country\_of\_birth + spedflag + TAG\_participant + Low\_SES + lunch\_status + Primary\_disability + dyslexia + atrisk + homeless , data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)}
\]

\[
\text{imp.data$match.weight[in.imp] \leftarrow m.out$weights}
\]

\[
\text{summary(m.out)}
\]

\[
\text{write.csv(imp.data, file =...
```r
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status", 
"Primary_disability", "dyslexia", "atrisk", "homeless", "AgeinDays", "grade"),
new = c("Race", "Ethnicity", "Sex", "Limited English Proficiency", "Primary 
Language", "Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low 
SES Status", "Lunch Status", "Disability Status", "Dyslexia", "At-risk Student", "Homelessness", 
"Grade"))

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp =
".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs =
TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
```

---

125
## Reading ATT estimation for the 100 imputed datasets

```r
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")
```

### Reading Sensitivity Test

```r
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
pens(tr1, tr1, Gamma=1.7, GammaInc=0.5)
```

### GRADE 2 SUPERA READING

```r
imp.data <- read.csv("C:/Users/>>>>/Desktop/supera02.csv")
View(imp.data)
```

### Full Propensity scores Estimation and Matching

```r
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
                        student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
                        lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,],
                        family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
                   country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
                   Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = "full",
                   ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file =
  "C:/Users/>>>>/Desktop/supera01full.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
Primary_disability + dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, get0ption("digits" - 3)))
## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG Participant", "Low SES", "lunch_status",
"Primary disability", "dyslexia", "atrisk", "homeless", "grade"),

## Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)

mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100`)

## Reading ATT estimation for the 100 imputed datasets
Y <- (imp.data$READ_SCORE)
z.att2.mi <- zelig(READ_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att2.mi, treatment = "HIPPY")
qi1 <- get_qi(z.att2.mi, qi = "ATT", xvalue = "TE")
hist(qi1, main = NULL, xlab ="Averege Treatment Effect of HIPPY on the Treated, Optimal Full Matching, Supera READING, Grade 2")
combine_coef_se(z.att2.mi)

# Sensitivity Analysis
Y <- imp.data$READ_SCORE
tr <- imp.data$HIPPY
psens(Y[tr =="0"], Y[tr =="1"], Gamma=1.7, GammaInc=0.5)
## Optimal Propensity scores Estimation and Matching

```r
imp.data <- read.csv("C:/Users/.../Desktop/supera02.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
    in.imp <- imp.data$.imp == i
    imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], family="binomial")$fitted.values
    m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)
    imp.data$match.weight[in.imp] <- m.out$weights
}

summary(m.out)
write.csv(imp.data, file = "C:/Users/.../Desktop/supera02optimal.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:

## Using alternate variable names

##Love Plot of Propensity Score Matching:
```
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)

##list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100`)

## Reading ATT estimation for the 100 imputed datasets
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")
# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, Gammalnc=0.5)

###GRADE 3 STAAR
## Full Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/staar03.csv")
View(imp.data)

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {
  in.imp <- imp.data$imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data =
imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +

Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,]
method = "full", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)

write.csv(imp.data, file =
"C:/Users/>>>>/Desktop/staar03full.csv")
imp.data[1:10,]

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

### visual inspection:
### Using alternate variable names

### Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,

130
## MATH ATT estimation for the 100 imputed datasets

z.att3.mi <- zelig(MATH_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att3.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att3.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR MATH, Grade 3")
combine_coef_se(z.att3.mi)

# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)

## Reading ATT estimation for the 100 imputed datasets

Y1 <- (imp.data$READ_SCORE)

z.att3.mi <- zelig(READ_SCORE ~ HIPPY, model = "ls", data = mi.out)
ATT(z.att3.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att3.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR READING, Grade 3")

# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching

imp.data <- read.csv("C:/Users/>>>>/Desktop/staar03.csv")
View(imp.data)

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
in.imp <- imp.data$.imp == i

imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], family="binomial")$fitted.values

m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp],
method = "optimal", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights }

summary(m.out)

write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar03optimal.csv")

imp.data[1:10,]

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless + AgeinDays, s.d.denom = "treated", data = imp.data, weights=
"match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)

print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names

v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status",
"Primary_disability", "dyslexia", "atrisk", "homeless", "AgeinDays", "grade"),

"Age in Days", "Grade"))

## Love Plot of Propensity Score Matching:

love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat =
"mean.diff", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits =
c(0, 1))

# Balance Plot of Propensity Score Matching:

bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")

# Define the multiple imputed data

m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)

## List of multiple imputed datasets.

miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,

132
## MATH ATT Estimation Multiple Imputation Welch two-sample t-test

```r
y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight")
```

# Math Sensitivity Test

```r
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)
```

## Reading ATT estimation for the 100 imputed datasets

```r
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")
```

# Reading Sensitivity Test

```r
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)
```

###GRADE 4 STAAR

```r
imp.data <- read.csv("C:/Users/>>>>/Desktop/staar04.csv")
View(imp.data)
```

---

```r
## Full Propensity scores Estimation and Matching
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {
  in.imp <- imp.data$imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
  student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
  lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data =
  imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
  country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
  ...)```

---

133
Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,],
method = "full", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)

write.csv(imp.data, file =
   "C:/Users/>>>>/Desktop/staar04full.csv")
imp.data[1:10,]

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAGParticipant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)

bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)

print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAGParticipant", "Low_SES", "lunch_status",
"Primary_disability", "dyslexia", "atrisk", "homeless", "grade"),

## Love Plot of Propensity Score Matching:
love.plot(bal.tab(m.out, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))
## MATH ATT estimation for the 100 imputed datasets
z.att4.mi <- zelig(MATH_SCORE ~ HIPPY, data = mi.out, model = "ls")
ATT(z.att4.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att4.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR MATH, Grade 4")
combine_coef_se(z.att4.mi)

# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)
## Reading ATT estimation for the 100 imputed datasets
Y1 <- (imp.data$READ_SCORE)
Y1 <- (imp.data$READ_SCORE)
z.att4.mi <- zelig(READ_SCORE ~ HIPPY, model = "ls", data = mi.out)
ATT(z.att4.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att4.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR READING, Grade 4")

# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)
## Optimal Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/staar04.csv")
View(imp.data)

imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {
in.imp <- imp.data$imp == i
imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp], family="binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,]
method = "optimal", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights }
summary(m.out)

write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar04optimal.csv")
imp.data[1:10,]

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless + AgeinDays, s.d.denom = "treated", data = imp.data, weights =
"match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status",
"Primary_disability", "dyslexia", "atrisk", "homeless", "AgeinDays", "grade"),
"Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low SES Status",
"Lunch Status", "Disability Stat", "Dyslexia", "At-risk Student", "Homelessness",
"Age in Days", "Grade"))

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat =
"mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits =
c(0, 1))

#Balance Plot of Propensity Score Matching:

bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")

#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)

##list of multiple imputed datasets.
miData <- to_zelig_mi(m.data$1`, m.data$2`, m.data$3`, m.data$4`, m.data$5`, m.data$6`,
m.data$7`, m.data$8`, m.data$9`, m.data$10`, m.data$11`, m.data$12`, m.data$13`,
m.data$14`, m.data$15`, m.data$16`, m.data$17`, m.data$18`, m.data$19`, m.data$20`,
```
m.data$'21}', m.data$'22}', m.data$'23}', m.data$'24}', m.data$'25}', m.data$'26}', m.data$'27}', m.data$'28}', m.data$'29}', m.data$'30}', m.data$'31}', m.data$'32}', m.data$'33}', m.data$'34}', m.data$'35}', m.data$'36}', m.data$'37}', m.data$'38}', m.data$'39}', m.data$'40}', m.data$'41}', m.data$'42}', m.data$'43}', m.data$'44}', m.data$'45}', m.data$'46}', m.data$'47}', m.data$'48}', m.data$'49}', m.data$'50}', m.data$'51}', m.data$'52}', m.data$'53}', m.data$'54}', m.data$'55}', m.data$'56}', m.data$'57}', m.data$'58}', m.data$'59}', m.data$'60}', m.data$'61}', m.data$'62}', m.data$'63}', m.data$'64}', m.data$'65}', m.data$'66}', m.data$'67}', m.data$'68}', m.data$'69}', m.data$'70}', m.data$'71}', m.data$'72}', m.data$'73}', m.data$'74}', m.data$'75}', m.data$'76}', m.data$'77}', m.data$'78}', m.data$'79}', m.data$'80}', m.data$'81}', m.data$'82}', m.data$'83}', m.data$'84}', m.data$'85}', m.data$'86}', m.data$'87}', m.data$'88}', m.data$'89}', m.data$'90}', m.data$'91}', m.data$'92}', m.data$'93}', m.data$'94}', m.data$'95}', m.data$'96}', m.data$'97}', m.data$'98}', m.data$'99}', m.data$'100}')

## MATH ATT Estimation Multiple Imputation Welch two-sample t-test

\[
y \leftarrow \text{miData$match.weight} \\
x \leftarrow \text{miData$MATH\_SCORE} \\
\text{mi.t.test(miData, x = "MATH\_SCORE", y = "match.weight")}
\]

# Math Sensitivity Test

\[
tr \leftarrow \text{imp.data$match.weight == "1"} \\
ctr \leftarrow \text{imp.data$match.weight == "0"} \\
\text{psens(ctr, tr, Gamma=1.7, GammaInc=0.5)}
\]

## Reading ATT estimation for the 100 imputed datasets

\[
y \leftarrow \text{miData$match.weight} \\
x \leftarrow \text{miData$READ\_SCORE} \\
\text{mi.t.test(miData, x = "READ\_SCORE", y = "match.weight")}
\]

# Reading Sensitivity Test

\[
tr1 \leftarrow \text{imp.data$match.weight == "1"} \\
ctr1 \leftarrow \text{imp.data$match.weight == "0"} \\
\text{psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)}
\]

###GRADE 5 STAAR

## Full Propensity scores Estimation and Matching

\[
\text{imp.data} \leftarrow \text{read.csv("C:/Users/>>>>/Desktop/staar05.csv")} \\
\text{View(imp.data)} \\
\text{imp.data$sps} \leftarrow \text{imp.data$match.weight} \leftarrow \text{rep(0, nrow(imp.data))} \\
\text{for (i in unique(imp.data$imp))} \{
\text{in.imp} \leftarrow \text{imp.data$imp == i} \\
\text{imp.data$sps[in.imp]} \leftarrow \text{glm(HIPPY ~ race + fedEthnicity + Sex + LEP\_program + student\_primary\_lang + country\_of\_birth + spedflag + TAG\_participant + Low\_SES + lunch\_status + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], family="binomial")$fitted.values}
\]
```
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
   country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk +
   homeless + AgeinDays, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar05full.csv")
View(imp.data)
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
   country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk +
   homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp =
   ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
   s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
   TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
   "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
   "country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status", "dyslexia",
   "atrisk", "homeless", "AgeinDays", "grade"),

## Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
   country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk +
   homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp =
   ".imp"), which.imp = 1, var.order = "unadjusted", threshold = .2, stat = "mean.diffs", abs =
   TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Balance Plot of Propensity Score Matching:
bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$imp))
View(m.data)
mi.out <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
   m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
   m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
   m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
   m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
138
## MATH ATT estimation for the 100 imputed datasets

```
z.att5.mi <- zelig(MATH_SCORE~ HIPPY, data = mi.out, model = "ls")
ATT(z.att5.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att5.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR MATH, Grade 5")
combine_coef_se(z.att5.mi)
```

# Math Sensitivity Test

```
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)
```

## Reading ATT estimation for the 100 imputed datasets

```
Y1 <- (imp.data$READ_SCORE)
z.att5.mi <- zelig(READ_SCORE ~ HIPPY, model = "ls", data = mi.out)
ATT(z.att5.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att5.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR READING, Grade 5")
```

# Reading Sensitivity Test

```
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)
```

###GRADE 5 STAAR

## Optimal Propensity scores Estimation and Matching

```
imp.data <- read.csv("C:/Users/\\\\\\\\Desktop/staar05.csv")
View(imp.data)
```

```
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {
in.imp <- imp.data$imp == i
```
imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], family = "binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights }
summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar05optimal.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk + homeless + AgeinDays, s.d.denom = "treated", data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continuous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names

#Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .2, stat = "mean.diff", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Balance Plot of Propensity Score Matching:
bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")

#Define the multiple imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
## list of multiple imputed datasets.

miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
                      m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
                      m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
                      m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
                      m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
                      m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
                      m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
                      m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
                      m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
                      m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
                      m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
                      m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
                      m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
                      m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
                      m.data$`98`, m.data$`99`, m.data$`100`)

## MATH ATT Estimation Multiple Imputation Welch two-sample t-test

y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight")

# Math Sensitivity Test

tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)

## Reading ATT estimation for the 100 imputed datasets

y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight")

# Reading Sensitivity Test

tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

### GRADE 6 STAAR

## Full Propensity scores Estimation and Matching

imp.data <- read.csv("C:/Users/username/Desktop/staar06.csv")
View(imp.data)

imp.data$sps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$$.imp)) {
  in.imp <- imp.data$$.imp == i
}
imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], family="binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar06full.csv")
View(imp.data)
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continuos = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names

## Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 1, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Balance Plot of Propensity Score Matching:
bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$imp))
View(m.data)

miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100`)

##MATH ATT estimation for the 100 imputed datasets
z.att6.mi <- zelig(MATH_SCORE~ HIPPY, data = mi.out, model = "ls")
ATT(z.att6.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att6.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR MATH, Grade 6")
combine_coef_se(z.att6.mi)

# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)

## Reading ATT estimation for the 100 imputed datasets
Y1 <- (imp.data$READ_SCORE)
z.att6.mi <- zelig(READ_SCORE ~ HIPPY, model = "ls", data = mi.out)
ATT(z.att6.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att6.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR READING, Grade 6")
combine_coef_se(z.att6.mi)

# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching for the 100 imputed datasets
imp.data <- read.csv("C:/Users/username/Desktop/staar06.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
  student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
  Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp],
  family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
  country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
  atrisk + homeless + AgeinDays, data = imp.data[in.imp], method = "optimal", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights }
summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar06optimal.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
atrisk + homeless + AgeinDays, s.d.denom = "treated", data = imp.data, weights =
"match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))
## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "Primary_disability",
"dyslexia", "atrisk", "homeless", "AgeinDays", "grade"),
   new = c("Race", "Ethnicity", "Sex", "Limited English Proficiency", "Primary 
Language", "Country of Birth", "Special Education Status", "Talented and Gifted Status", "Low 
"Grade"))

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat =
"mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits =
c(0, 1))
#Balance Plot of Propensity Score Matching:
bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")
# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100`)  

## MATH ATT Estimation Multiple Imputation Welch two-sample t-test
y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight", var.equal = TRUE)

# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
pseens(ctr, tr, Gamma=1.7, GammaInc=0.5)

## Reading ATT estimation for the 100 imputed datasets
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight", var.equal = TRUE)

# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
pseens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

### GRADE 8 STAAR
## Full Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/username/Desktop/staar07.csv")
View(imp.data)
imp.data$SpS <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$imp)) {

145
in.imp <- imp.data$.imp == i
imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data =
imp.data[in.imp[,], family="binomial"]$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status +
Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp[,],
method = "full", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights
}
}
}
summary(m.out)
write.csv(imp.data, file = "C:/Users/username/Desktop/staar07full.csv")
View(imp.data)
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability +
dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab =
TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
v <- data.frame(old = c("race", "fedEthnicity", "Sex", "LEP_program", "student_primary_lang ",
"country_of_birth", "spedflag", "TAG_participant", "Low_SES", "lunch_status",
"Primary_disability", "dyslexia", "atrisk", "homeless", "AgeinDays", "grade"),
"Lunch Status", "Disability Status", "Dyslexia", "At-risk Student", "Homelessness", 
"Age in Days", "Grade"))

##Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia +
atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp"), which.imp = 1, var.order = "unadjusted", threshold = .1, stat =
"mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits =
c(0, 1))

#Balance Plot of Propensity Score Matching:
bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")
# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
  m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
  m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
  m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
  m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
  m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
  m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
  m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
  m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
  m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
  m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
  m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
  m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
  m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
  m.data$`98`, m.data$`99`, m.data$`100`)
## MATH ATT estimation for the 100 imputed datasets
z.att7.mi <- zelig(MATH_SCORE~ HIPPY, data = mi.out, model = "ls")
ATT(z.att7.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att7.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab ="Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR MATH, Grade 7")
combine_coef_se(z.att7.mi)
# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)

## Reading ATT estimation for the 100 imputed datasets
Y1 <- (imp.data$READ_SCORE)
z.att7.mi <- zelig(READ_SCORE ~ HIPPY, model = "ls", data = mi.out)
ATT(z.att7.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att7.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab ="Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR READING, Grade 7")

# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/username/Desktop/staar07.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data[in.imp,], method = "optimal", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights }
summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar07optimal.csv")
imp.data[1:10,]
bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", s.d.denom = "treated", data = imp.data, imp.summary = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.ratio = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.threshold = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names

## Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + Primary_disability + dyslexia + atrisk + homeless + AgeinDays, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diff", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`,
m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`,
m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`,
m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`,
m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`,
m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`,
m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`,
m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`,
m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`,
m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`,
m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`,
m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`,
m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100`)

## MATH ATT Estimation Multiple Imputation Welch two-sample t-test
y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight", var.equal = TRUE)

# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)

## Full Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/staar08.csv")
View(imp.data)

## Full Propensity scores Estimation and Matching
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], family="binomial")$fitted.values
m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,], method = "full", ratio = 1, replace = TRUE)
imp.data$match.weight[in.imp] <- m.out$weights

summary(m.out)
write.csv(imp.data, file = "C:/Users/username/Desktop/staar08full.csv")
View(imp.data)

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continous = "std", disp.v.ratio = TRUE, un = TRUE, s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2, disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means = "as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, get0ption("digits" - 3)))

## visual inspection:
## Using alternate variable names
## Using alternate variable names

###Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + lunch_status+ Low_SES + Primary_disability+ dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

#Balance Plot of Propensity Score Matching:
bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")
# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100`)

## MATH ATT estimation for the 100 imputed datasets
z.att8.mi <- zelig(MATH_SCORE ~ HIPPY, data = miData, model = "ls")
ATT(z.att8.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att8.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Averge Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR MATH, Grade 8")
combine_coef_se(z.att8.mi)

#### simulation of ATT estimation for the 100 imputed datasets
z.out8.mi <- zelig(MATH_SCORE ~ HIPPY, data = mi.out, model = "ls")
x.out8 <- setx(z.out8.mi, HIPPY = 0)
x1.out <- setx(z.out8.mi, HIPPY = 1)
s.out8 <- sim(z.out8.mi, x = x.out8, x1 = x1.out)
summary(s.out8)
combine_coef_se(z.out8.mi)

# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)

## Reading ATT estimation for the 100 imputed datasets
Y1 <- (imp.data$READ_SCORE)
z.att8.mi <- zelig(READ_SCORE ~ HIPPY, model = "ls", data = miData)
ATT(z.att8.mi, treatment = "HIPPY")
qi0 <- get_qi(z.att8.mi, qi = "ATT", xvalue = "TE")
hist(qi0, main = NULL, xlab = "Average Treatment Effect of HIPPY on the Treated, Optimal Full Matching, STAAR READING, Grade 8")

# Reading Sensitivity Test
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)

## Optimal Propensity scores Estimation and Matching
imp.data <- read.csv("C:/Users/>>>>/Desktop/staar08.csv")
View(imp.data)
imp.data$ps <- imp.data$match.weight <- rep(0, nrow(imp.data))
for (i in unique(imp.data$.imp)) {
  in.imp <- imp.data$.imp == i
  imp.data$ps[in.imp] <- glm(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
                               student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
                               lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,],
                               family="binomial")$fitted.values
  m.out <- matchit(HIPPY ~ race + fedEthnicity + Sex + LEP_program +
                   student_primary_lang + country_of_birth + spedflag + TAG_participant + Low_SES +
                   lunch_status + Primary_disability + dyslexia + atrisk + homeless, data = imp.data[in.imp,],
                   method = "optimal", ratio = 1, replace = TRUE)
  imp.data$match.weight[in.imp] <- m.out$weights
}
summary(m.out)
write.csv(imp.data, file = "C:/Users/>>>>/Desktop/staar08optimal.csv")
imp.data[1:10,]

bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang +
country_of_birth + spedflag + TAG_participant + Low_SES + lunch_status + dyslexia + atrisk +
homeless, s.d.denom = "treated", data = imp.data, weights = "match.weight", method =
"matching", imp = ".imp", imp.summary = TRUE)
bal.tab(m.out, int = TRUE, binary = "raw", continuous = "std", disp.v.ratio = TRUE, un = TRUE,
s.d.denom = c("treated", "control", "pooled"), m.threshold = .1, v.threshold = 2,
        disp.bal.tab = TRUE, disp.means = TRUE)
print(m.out, disp.m.threshold = "as.is", disp.v.threshold = "as.is", un = "as.is", disp.means =
"as.is", disp.bal.tab = "as.is", disp.v.ratio = "as.is", digits = max(3, getOption("digits" - 3)))

## visual inspection:
## Using alternate variable names
## Love Plot of Propensity Score Matching:
love.plot(bal.tab(HIPPY ~ race + fedEthnicity + Sex + LEP_program + student_primary_lang + country_of_birth + spedflag + TAG_participant + lunch_status + Low_SES + Primary_disability + dyslexia + atrisk + homeless, data = imp.data, weights = "match.weight", method = "matching", imp = ".imp"), which.imp = 100, var.order = "unadjusted", threshold = .1, stat = "mean.diffs", abs = TRUE, int = TRUE, line = TRUE, var.names = v, no.missing = TRUE, limits = c(0, 1))

# Balance Plot of Propensity Score Matching:
bal.plot(m.out, var.name = "distance", mirror = TRUE, type = "histogram")

# Define the multiply imputed data
m.data <- with(match.data(m.out), split(imp.data, imp.data$.imp))
View(m.data)
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`, m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`, m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`, m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`, m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`, m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`, m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`, m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`, m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`, m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`, m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`, m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`, m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`, m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`, m.data$`98`, m.data$`99`, m.data$`100")

## MATH ATT Estimation Multiple Imputation Welch two-sample t-test
y <- miData$match.weight
x <- miData$MATH_SCORE
mi.t.test(miData, x = "MATH_SCORE", y = "match.weight", var.equal = TRUE)

# Math Sensitivity Test
tr <- imp.data$match.weight == "1"
ctr <- imp.data$match.weight == "0"
psens(ctr, tr, Gamma=1.7, GammaInc=0.5)
## Reading ATT estimation for the 100 imputed datasets

```r
y <- miData$match.weight
x <- miData$READ_SCORE
mi.t.test(miData, x = "READ_SCORE", y = "match.weight", var.equal = TRUE)
```

# Reading Sensitivity Test
```r
tr1 <- imp.data$match.weight == "1"
ctr1 <- imp.data$match.weight == "0"
psens(ctr1, tr1, Gamma=1.7, GammaInc=0.5)
```
APPENDIX C

R-CODE FOR INDEPENDENT SAMPLE t-TEST ON THE Z-SCORES
### Independent Sample t-test on the Z-scores for Each Assessment

```r
install.packages("Zelig")
library(Zelig)
install.packages("MKmisc")
library("MKmisc")
###ISIP
imp.data <- read.csv("C:/Users/username/Desktop/isip00.csv")
View(imp.data)
# Define the multiply imputed data
m.data <- with(imp.data, split(imp.data, imp.data$.imp))
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100`)
# independent sample t-test
y <- miData$HIPPY
x <- miData$ZOverall_SCORE
mi.t.test(miData, x = "ZOverall_SCORE", y = "HIPPY", var.equal = TRUE)
###TerraNova Math
imp.data <- read.csv("C:/Users/username/Desktop/terranovamath0-2.csv")
View(imp.data)
# Define the multiply imputed data
m.data <- with(imp.data, split(imp.data, imp.data$.imp))
View(m.data)
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100`)
# independent sample t-test
y <- miData$HIPPY
x <- miData$ZOverall_SCORE
mi.t.test(miData, x = "ZOverall_SCORE", y = "HIPPY", var.equal = TRUE)
```
# independent sample t-test

```r
y <- miData$HIPPY
x <- miData$ZMATH_SCORE
mi.t.test(miData, x = "ZMATH_SCORE", y = "HIPPY", var.equal = TRUE)
```

### TerraNova Reading

```r
imp.data <- read.csv("C:/Users/>>>>/Desktop/terranovaread.0-2.csv")
View(imp.data)
```

# Define the multiply imputed data

```r
m.data <- with(imp.data, split(imp.data, imp.data$.imp))
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100`)
```

# independent sample t-test

```r
y <- miData$HIPPY
x <- miData$ZREAD_SCORE
tt <- mi.t.test(miData, x = "ZREAD_SCORE", y = "HIPPY", var.equal = TRUE)
```

### Supera

```r
imp.data <- read.csv("C:/Users/>>>>/Desktop/supera0-2.csv")
View(imp.data)
```

# Define the multiply imputed data

```r
m.data <- with(imp.data, split(imp.data, imp.data$.imp))
miData <- to_zelig_mi(m.data$`1`, m.data$`2`, m.data$`3`, m.data$`4`, m.data$`5`, m.data$`6`,
m.data$`7`, m.data$`8`, m.data$`9`, m.data$`10`, m.data$`11`, m.data$`12`, m.data$`13`,
m.data$`14`, m.data$`15`, m.data$`16`, m.data$`17`, m.data$`18`, m.data$`19`, m.data$`20`,
m.data$`21`, m.data$`22`, m.data$`23`, m.data$`24`, m.data$`25`, m.data$`26`, m.data$`27`,
m.data$`28`, m.data$`29`, m.data$`30`, m.data$`31`, m.data$`32`, m.data$`33`, m.data$`34`,
m.data$`35`, m.data$`36`, m.data$`37`, m.data$`38`, m.data$`39`, m.data$`40`, m.data$`41`,
m.data$`42`, m.data$`43`, m.data$`44`, m.data$`45`, m.data$`46`, m.data$`47`, m.data$`48`,
m.data$`49`, m.data$`50`, m.data$`51`, m.data$`52`, m.data$`53`, m.data$`54`, m.data$`55`,
m.data$`56`, m.data$`57`, m.data$`58`, m.data$`59`, m.data$`60`, m.data$`61`, m.data$`62`,
m.data$`63`, m.data$`64`, m.data$`65`, m.data$`66`, m.data$`67`, m.data$`68`, m.data$`69`,
m.data$`70`, m.data$`71`, m.data$`72`, m.data$`73`, m.data$`74`, m.data$`75`, m.data$`76`,
m.data$`77`, m.data$`78`, m.data$`79`, m.data$`80`, m.data$`81`, m.data$`82`, m.data$`83`,
m.data$`84`, m.data$`85`, m.data$`86`, m.data$`87`, m.data$`88`, m.data$`89`, m.data$`90`,
m.data$`91`, m.data$`92`, m.data$`93`, m.data$`94`, m.data$`95`, m.data$`96`, m.data$`97`,
m.data$`98`, m.data$`99`, m.data$`100`)
```
```
m.data$'49', m.data$'50', m.data$'51', m.data$'52', m.data$'53', m.data$'54', m.data$'55',
  m.data$'56', m.data$'57', m.data$'58', m.data$'59', m.data$'60', m.data$'61', m.data$'62',
  m.data$'63', m.data$'64', m.data$'65', m.data$'66', m.data$'67', m.data$'68', m.data$'69',
  m.data$'70', m.data$'71', m.data$'72', m.data$'73', m.data$'74', m.data$'75', m.data$'76',
  m.data$'77', m.data$'78', m.data$'79', m.data$'80', m.data$'81', m.data$'82', m.data$'83',
  m.data$'84', m.data$'85', m.data$'86', m.data$'87', m.data$'88', m.data$'89', m.data$'90',
  m.data$'91', m.data$'92', m.data$'93', m.data$'94', m.data$'95', m.data$'96', m.data$'97',
  m.data$'98', m.data$'99', m.data$'100')
# independent sample t-test
y <- miData$'HIPPY'
x <- miData$'ZREAD_SCORE'
mi.t.test(miData, x = "ZREAD_SCORE", y = "HIPPY", var.equal = TRUE)
### STAAR
# Define the multiply imputed data
m.data <- with(imp.data, split(imp.data, imp.data$.imp))
miData <- to_zelig_mi(m.data$'1', m.data$'2', m.data$'3', m.data$'4', m.data$'5', m.data$'6',
  m.data$'7', m.data$'8', m.data$'9', m.data$'10', m.data$'11', m.data$'12', m.data$'13',
  m.data$'14', m.data$'15', m.data$'16', m.data$'17', m.data$'18', m.data$'19', m.data$'20',
  m.data$'21', m.data$'22', m.data$'23', m.data$'24', m.data$'25', m.data$'26',
  m.data$'27', m.data$'28', m.data$'29', m.data$'30', m.data$'31', m.data$'32', m.data$'33',
  m.data$'34', m.data$'35', m.data$'36', m.data$'37', m.data$'38', m.data$'39', m.data$'40',
  m.data$'41', m.data$'42', m.data$'43', m.data$'44', m.data$'45', m.data$'46', m.data$'47',
  m.data$'48', m.data$'49', m.data$'50', m.data$'51', m.data$'52', m.data$'53', m.data$'54',
  m.data$'55', m.data$'56', m.data$'57', m.data$'58', m.data$'59', m.data$'60', m.data$'61',
  m.data$'62', m.data$'63', m.data$'64', m.data$'65', m.data$'66', m.data$'67', m.data$'68',
  m.data$'69', m.data$'70', m.data$'71', m.data$'72', m.data$'73', m.data$'74', m.data$'75',
  m.data$'76', m.data$'77', m.data$'78', m.data$'79', m.data$'80', m.data$'81', m.data$'82',
  m.data$'83', m.data$'84', m.data$'85', m.data$'86', m.data$'87', m.data$'88', m.data$'89',
  m.data$'90', m.data$'91', m.data$'92', m.data$'93', m.data$'94', m.data$'95', m.data$'96',
  m.data$'97', m.data$'98', m.data$'99', m.data$'100')
# independent sample t-test for Math
y <- miData$'HIPPY'
x <- miData$'ZMATH_SCORE'
mi.t.test(miData, x = "ZMATH_SCORE", y = "HIPPY", var.equal = TRUE)
# independent sample t-test for Reading
y <- miData$'HIPPY'
x <- miData$'ZREAD_SCORE'
mi.t.test(miData, x = "ZREAD_SCORE", y = "HIPPY", var.equal = TRUE)
APPENDIX D

IRB APPROVAL
September 5, 2018

Dr. Robin Henson
Student Investigator: Amal Abdulaziz
Department of Educational Psychology
University of North Texas
RE: Human Subjects Application No. 18-128

Dear Dr. Henson:

In accordance with 45 CFR Part 46 Section 46.101, your study titled “Impact Evaluation of Texas HIPFY Program” has been determined to qualify for an exemption from further review by the UNT Institutional Review Board (IRB).

No changes may be made to your study’s procedures or forms without prior written approval from the UNT IRB. Please contact The Office of Research Integrity and Compliance at 940-565-4643, if you wish to make any such changes. Any changes to your procedures or forms after 3 years will require completion of a new IRB application.

We wish you success with your study.

Sincerely,

Shelley Riggs, Ph.D.
Professor
Chair, Institutional Review Board

Sfjm


