Blondin, C. A., Voils, K., Galyon, C. E., & Williams, R. L. (2015). Selected application of response-to-intervention principles in college courses: Possibilities and limitations. *Journal on Excellence in College Teaching, 26*(2), 61-83.

Selected Application of Response-to-Intervention Principles in College Courses: Possibilities and Limitations

Carolyn A. Blondin
Kyle Voils
Charles E. Galyon
Robert L. Williams
The University of Tennessee, Knoxville

Concepts from the Response-to-Intervention (RTI) Model were used to promote a successful course outcome for students at risk for making low grades in an entry-level college course. The first exam served as a universal screener to identify students who could potentially benefit from RTI assistance. The researchers developed a tiered coaching arrangement targeting homework accuracy and exam scores. Homework accuracy was used as a progress-monitoring measure. A multiple-baseline design across participants showed that homework accuracy increased from baseline to treatment phases. Exam gaps between participant exam scores and class averages decreased following treatment implementation for a majority of participants.

Because performance in large entry-level undergraduate courses is often assessed primarily through multiple-choice examinations, poor performance on early multiple-choice exams may signal a need for an intervention to prevent a pattern of poor test performance and possible failure in such courses. Unfortunately, poor performance in large entry-level courses constitutes one of the major predictors of students' eventually dropping out of college (Reason, 2009). Thus, identifying contributors to performance on multiple-choice exams is a major first step in improving

exam scores in large entry-level courses. The findings of Galyon, Blondin, Forbes, and Williams (2013) suggest that critical thinking and homework accuracy are the most consistent academic predictors of exam performance in courses similar to the one targeted in this study. Consequently, interventions designed to improve exam performance should target one or both of these predictors.

Response to Intervention

Nationwide, primary and secondary schools have implemented school-wide Response to Intervention (RTI) models in order to detect and remediate early difficulties in academic areas such as math, writing, and reading (Burns, Appleton, & Stehouwer, 2005). Reutebuch (2008) described RTI as a framework that provides high-quality instruction to all students and tailors supplemental interventions to meet the needs of students doing poorly under class-wide instruction. RTI models typically adhere to the following guidelines: (a) All students receive research-based instruction in the general education classroom, (b) universal screenings of all students are conducted to assess the attainment and maintenance of an academic skill, (c) struggling students receive early intervention that targets their academic weaknesses, (d) progress is monitored frequently for students receiving extra instruction to determine responsiveness to intervention, and (e) interventions are tier-based and intensified in terms of time or resources for non-responding students (Jimerson, Burns, & VanDerHeyden, 2007; Mellard, McKnight, & Woods, 2009; Reutebuch, 2008). Despite RTI's widespread use at the elementary level and somewhat limited use at the secondary level, its use at the college level has been virtually non-existent.

We anticipate that certain features of the RTI model could be a beneficial framework for identifying college students struggling in entry-level courses and then providing interventions early enough to increase their prospects for success in these courses. College professors who employ early evidenced-based assistance for struggling students will be likely to have more success in improving their individual academic outcomes than those who provide remedial assistance after repeated failure experiences in a course. Given the predictive potential of early exam performance (Blondin, 2012), we identified principles in the RTI model that might counter the lingering effects of low scores on early exams. Mellard (2005) highlighted the core features of RTI as research-based, class-wide instruction; universal screening; data-based small-group and individualized interventions; and continuous-progress monitoring, all of which we attempted with several students who did poorly on the initial exam in the large entry-level course we teach.

RTI is typically organized in three tiers that progress from large-group, to small-group, and eventually to individual instruction. A general first tier of RTI provides universal instruction and services to all the students. Fuchs, Fuchs, and Compton (2010) recommended that the core instructional program in the first tier include research-based teaching methods, differential instruction, classroom accommodations, and problem-solving strategies to address challenges related to academic motivation. RTI programs use universal screeners to identify students who are not responding to the large-group instruction. Selection decisions are made by comparing students' performance on a curriculum measure to that of peers. Identified students are subsequently provided more intensive small-group instruction (Tier 2). Progress-monitoring tools are used regularly to determine progress of students in Tier 2 interventions (Mellard, 2005). If students fail to make sufficient progress after a predetermined period of engagement in small-group instruction, they are transitioned to more intensive and individualized coaching (Tier 3). Tier 3 interventions commonly differ from Tier 2 interventions in that Tier 3 interventions are often of longer duration (for example, 45-minute sessions in Tier 3 vs. 30-minute sessions in Tier 2), occur more frequently (for example, five times a week in Tier 3 vs. three times a week in Tier 2), and have a lower student-teacher ratio (for example, one-on-one instruction in Tier 3 vs. small-group instruction in Tier 2) (Mellard, McKnight, & Jordan, 2010). The instructional content for Tier 3 is typically individualized and is based on a student's progress monitoring data.

Because of the flexibility required in scheduling the tiers, the RTI model is typically applied on a whole-school basis. In addition to scheduling large-group instruction, schools must build time into the regular schedule for small-group and individualized instruction. However, there is some precedent for applying RTI on a class-wide rather than a school-wide basis (Ardoin, Witt, Connell, & Koenig, 2005; Tackett, Roberts, Baker, & Scammaca, 2009). Reasons that school officials have elected to implement pilot programs on a smaller scale include determining what works, what changes are needed, and what resources are necessary before making a school-wide commitment (Tackett et al., 2009).

Although the majority of studies on RTI have been conducted at the primary school level (Burns et al., 2005), increased attention is now being given to applying RTI at the secondary school level (Duffy, 2007; Ehren & Deshler, 2010). Researchers agree that RTI decisions at the secondary level should be based on multiple sources of empirical data, include tiered supplemental help in both skill and content mastery, and use collaborative problem-analysis in determining progress and interventions (Burns,

2008; Shinn, 2008). Students in middle and high school may have complex academic concerns that transcend problems with basic academic skills. Researchers have recommended that content learning be improved by focusing on both basic and high-level literacy skills, such as proficiency in listening, reading, and writing (Shinn, 2008). Modifications to interventions and changes in screening and progress monitoring measures appear warranted at the high school level (Ehren, 2008; Ehren & Deshler, 2010).

Reasons that RTI may have limited application at the college level likely include the multiplicity of course topics, diverse course structures, and minimal out-of-class interaction between students and professors. Because of the arrangement of college courses, time is generally not available within the conventional course for students to engage in Tier-2 and Tier-3 activities during class. Given the independence between most college courses, it is unlikely that RTI could be applied in a whole-school setting in college. Nonetheless, most of the RTI core principles identified by Mellard (2005) appear applicable within individual college courses (Jimerson et al., 2007; Mellard et al., 2009; Reutebuch, 2008). While widespread application of the total RTI model may not be feasible at the college level, some features of this model could be helpful to students struggling to make acceptable grades in large entry-level courses.

Framework for the Study

This study focused on improving overall exam and homework performance by using a tier-based intervention program that included early identification of low performers and use of progress monitoring tools to measure response to intervention. Tier 1 included a multiplicity of instructional arrangements that have contributed to success in the targeted course (Carroll, Williams, & Hautau, 2006; Foster et al., 2009; Galyon et al., 2013; Hautau et al., 2006; Krohn et al., 2010; Skinner, Williams, & Neddenriep, 2004; Turner & Williams, 2007; Williams, Oliver, & Stockdale, 2004). These procedures most notably included provision of a full set of instructor notes to all students at the beginning of the course plus study questions over the instructor notes for students to answer in writing prior to their being discussed in class and encountered on exams. All classes were characterized by instructor-led discussion of issues and concepts embedded in answers to the study questions.

We designed participant-selection procedures and implementation of out-of-class interventions for students who performed poorly on the first exam. At the conclusion of the Tier-1 activities, scores on the first unit exam were used as universal screeners to determine which students needed

assistance beyond that provided in regular class sessions. The universal criterion for needed assistance was a D or F on the initial exam. Tier-2's intervention focused on increasing test-taking and homework strategies by reviewing students' early exam performance and pinpointing areas to strengthen. If little to no improvement occurred in Tier 2, as determined by subsequent homework and exam performance, the intervention in Tier 3 was intensified to include several one-on-one sessions targeting students' overall understanding and critical application of key concepts outlined in the current course unit.

Thus, this study was designed to evaluate the possibility of implementing several components of RTI in a large undergraduate teacher-education course**.** We sought to determine what RTI components are feasible in the college setting, what components need modification to be implemented, and what limitations existed in the application of the RTI model at the collegiate level. Although Galyon et al. (2013) have determined homework accuracy and critical thinking to be strong predictors of exam success, critical thinking has not proven to be a highly modifiable predictor (Williams, Oliver, & Stockdale, 2004). Thus, we chose to target homework accuracy as a more modifiable avenue for improving exam performance.

Methods

Overall Student Sample

This study was conducted in three sections (Sections 1, 2, and 3) of an entry-level educational psychology course. The average class size was 54 students. Approximately 72% of the students were female. Two percent of students were freshmen, 47% were sophomores, 29% were juniors, 14% were seniors, 3% were graduate students, and 5% were non-degree-seeking or did not indicate an academic classification. The average reported GPA across sections was 3.22 (out of 4.0). Institutional Review Board approval and informed consent to participate were obtained prior to the commencement of this study.

Because critical thinking consistently has been found to be strongly related to exam performance (Wallace & Williams, 2003; Williams, Oliver, Allin, Win, & Booher, 2003; Williams et al., 2004; Williams & Stockdale, 2003), we assessed critical thinking at the beginning of the course and determined the extent to which the critical thinking (CT) scores of poor performers on the first exam differed from the CT scores of those who performed adequately on the exam. All students completed the *Watson-Glaser Critical Thinking Test-Form-S* (WGCTA, Watson & Glaser, 1994) at the outset

of the course. Percentile ranks on the WGCTA were established through comparison to a normative sample of college graduates. Across sections, the combined critical thinking mean was at the 22nd percentile and the median at the 15th percentile of the normative sample.

Selection of RTI Participants

Students who scored a 69% or lower (D or F) on the first exam—the study's universal screener—were invited to participate in the study. Across sections, 122 out of 164 students made a C or higher on the first exam, suggesting adequate performance under Tier-1 activities for a majority of students. Forty-two students (approximately 26%) made a D or F on the first exam and were invited to participate in the tier-level arrangement. Ten of the students contacted, all females, elected to participate in the study. Of the 10 initial participants, 7 remained with the full RTI services, while 3 eventually elected to discontinue intervention services rather than proceed to Tier 3. Three of the seven full participants experienced both Tier 2 and Tier 3. For confidentiality purposes, the names of intervention participants have been changed. Six of the original participants were sophomores, three were juniors, and one was non-degree-seeking. The demographic data indicated that participants had done relatively well in their previous courses (average GPA of 3.21), had low critical-thinking skills (M = 6th percentile), and attended class regularly during the course (average attendance rate of 97%).

RTI Intervention Tiers

Tier 1

The initial intervention tier included the regular classroom activities in the course, which were divided into five units (Units A-E) covering different dimensions of human development. The overall structure of the course was based on several evidence-based practices: providing instructor notes, homework questions over these notes, and class discussion of answers to the homework questions. Credit was given for students' attendance, completion of homework questions, participation in class discussions, and use of practice exams to prepare for unit exams. Each discussion day was audio-recorded and available for students to review. Prior to each exam, instructors posted two documents related to the unit: instructor answers to all of the homework questions and an exam-sources document that identified specific sections in the course documents pertaining to each exam item. Each unit included a 50-item multiple-choice exam covering

information presented during the unit. Each exam was worth up to 50 points, and the combination of all unit exams approximated 40% of total points in the course.

Tier 2

The primary researcher, who had taught the course for 3 years, provided Tier-2 activities to individuals making a D or F on the first unit exam. This tier commenced at differing times during the semester for students in different sections. Eligible students received notification from their teacher regarding the exam-improvement option. Implementation of all intervention procedures was based on each student's willingness to participate in the sessions. The primary researcher led small-group sessions, where she provided direct instruction on test-taking skills, tactics for reading exam questions, and connections between homework material and exam items. Students met on the basis of their availability, with attendance ranging from one to three individuals per session. Students were asked to bring course materials to the study sessions, including their instructor-notes handbook, exam-sources document, and instructor answers to the discussion questions. The primary researcher made available the exams and participants' previously used scan forms for the exams. Each meeting took place in a private office and lasted 30 minutes. One session was scheduled per unit per individual.

During Tier 2, the primary researcher asked the students to discuss what difficulties they experienced while taking the first test and how they had prepared for that exam. The primary researcher then reviewed frequently missed exam items, identifying any gaps between a student's knowledge of a particular concept and the answer(s) to exam item(s) targeting that concept. Using the exam-sources document and instructor answers to the discussion questions, the primary researcher demonstrated how the missed exam items corresponded to homework questions and instructor notes. Strategies for accurately answering exam questions were reviewed (for instance, eliminating unsupportable options; being cautious about extreme descriptions). The researcher also provided general tips for completing homework (for instance, answering all parts of a question, including all relevant concepts in their answers) and referred to resources available to help students construct answers to the homework questions.

Tier 3

If participants continued to show little or no improvement in exam scores and homework accuracy in Tier 2, they received individual instruction on improving the accuracy of their homework. Students were

asked to meet with the primary researcher twice during the Tier 3 unit for a 45-minute individual session. During these sessions, they discussed students' answers to selected homework questions, and the instructor pointed out incomplete or inaccurate answers. Students were encouraged to solicit instructor clarification regarding concepts related to homework questions. The goals of this intervention were to help students be more thorough and accurate in answering the homework questions and, as a result, gain a solid understanding of the material on which the exams were based.

Progress-Monitoring Measure

For several reasons, students' performance on the homework assignments was chosen to serve as the progress-monitoring tool in the study. Homework performance has been one of the better predictors of exam performance in the target course and occurs with sufficient frequency to provide both diagnostic and formative assessment (Galyon et al., 2013). The homework assignments addressed many of the issues embedded in the course exams, making improvement on the homework assignments a likely contributor to improvement on exam performance. Past research (Galyon et al., 2013) has yielded high inter-rater agreement in the scoring of homework performance (84%), and performance on homework assignment has been found to be a significant predictor of exam performance.

All students in the class were assigned daily homework on the four discussion days of each unit. Evaluating answers to the homework questions for the seven participants consisted of comparing the number of concepts accurately explained in a student's answer to the total number of concepts in the instructor's answer for the selected question (Homework-Accuracy Score). Inasmuch as Galyon and colleagues (2013) concluded that homework accuracy was a significant and primary predictor of exam performance, we expected the Daily Homework-Accuracy Score to be the best representation of content mastery. A high Homework-Accuracy Score indicated that the student's answer included a high percentage of accurate core concepts represented in the instructor's answer, while a low Homework-Accuracy Score indicated that very few core concepts were accurately reflected.

Assessment of homework accuracy followed procedures similar to those articulated by Galyon et al. (2013). Questions relating to the most exam items and covering the most unit concepts were selected for accuracy scoring. Progress-monitoring data were collected on each discussion day and consisted of the average percentage of Homework-Accuracy Scores

across targeted questions. Fifty-two questions were evaluated for homework accuracy per participant, with a range of 9 to 12 questions chosen per unit. The modal number of questions rated per day was two. A homework question omitted by a student was not included in the scoring.

Data-Based Decisions Regarding Tier Placement

Homework-Accuracy Scores were used to evaluate the effectiveness of the RTI intervention and provide a framework for decisions on student tier placement. The decision to keep a student at a current level of intervention, to cease providing supplemental intervention, or to move him or her to a more intensive intervention were contingent on specified criteria for the student's progress: (a) If the student had not made adequate improvement on homework accuracy (that is, the average increase on the Homework-Accuracy Score was lower than an increase of 10 percentage points from the student's homework average in the previous unit), then the student was placed in a more intensive tier; (b) if a student's Homework-Accuracy Score average was equal to or higher than a 10 percentage point increase from the homework-accuracy average of the previous unit, but the student still made a D or F on the current unit exam, the student remained in the current tier; (c) if the student achieved at least a 10 percentage point increase on homework-accuracy average from the previous unit and earned a C or higher on the next unit exam, the student met the criteria to return to Tier 1. Maintenance data on homework accuracy and exam performance were collected for any participant who graduated back to Tier 1. Criteria for tier placement following Tier 3 were similar to the decision-criteria for Tier 2: The student remained in Tier 3 if homework accuracy did not improve, returned to Tier 2 if homework accuracy improved but exam performance remained low, or graduated back to Tier 1 if both homework accuracy and exam performance improved.

Research Design

A multiple-baseline design across sections was used to evaluate the effectiveness of the interventions, as measured by homework accuracy. Performance on the Unit A exam was the only screener for participant eligibility. However, the point at which individuals were asked to participate differed depending on their course section: Students in Section 1 were invited to participate immediately after the first exam; Section 2 students were invited following the second exam; and Section 3 students were invited to participate immediately following the third exam. Thus,

Tier 2 and Tier 3 interventions were also introduced at different points in the three sections. Depending on the section, participants had between 5 and 12 data points measuring their homework accuracy during the baseline phase (see Figure 1). Treatment effects were determined by the percentage of data points in the RTI phase that exceeded the median data point in the baseline phase. Although the Unit A exam was the sole initial screener, baseline/Tier 1 exams scores were determined by averaging performance across exams prior to invitation to participate in the project (Section 1 Baseline included Unit A; Section 2 Baseline included Units A-B; Section 3 Baseline included Units A-C).

Inter-Observer Agreement

Homework questions targeted in this study had been previously selected by Galyon and colleagues (2013). Inter-rater agreement data for Homework-Accuracy Scores were calculated by procedures delineated by Galyon et al (2013). Prior to inter-rater data collection, the developer of the homework accuracy rubric taught grading methods to the primary researcher of this study and a comparison rater. In this study, participants' Homework-Accuracy Scores were evaluated for inter-rater agreement for approximately 27% of the questions selected. Fourteen out of the total of 52 questions per participant were evaluated for inter-rater agreement. Percent inter-observer agreement was calculated by dividing the number of evaluator agreements on homework answers by the total number of answers rated, multiplied by 100. Inter-observer agreement averaged 85% across the five course units.

Treatment Integrity

The primary researcher conducted a total of 18 intervention sessions (12 Tier-2 sessions and 6 Tier-3 sessions). To adhere to the standard treatment protocol, the primary researcher followed a procedural checklist when implementing interventions. Completed checklists indicated correct application of 100% of the steps involved in treatment implementation across all 18 sessions. The primary researcher audio-recorded the intervention sessions, and then an independent observer, a research assistant familiar with the course material, reviewed the recordings. He randomly picked six intervention sessions (three Tier-2 sessions and three Tier-3 sessions) and used procedural checklists to rate adherence to procedures. Checklists completed by the observer confirmed that the primary researcher implemented 100% of the steps correctly during the six recorded sessions.

Insert Figure 1 as a foldout,
sideways.

Results

Homework Improvement

Changes in participants' overall Daily Homework-Accuracy Scores were analyzed by computing the percentage of intervention data points exceeding the median of the baseline/Tier 1 (PEM; Ma, 2006). Due to the high sensitivity to potential outliers within the progress monitoring measure, a PEM index was selected rather than a percentage of non-overlapping data (PND) approach to assess treatment effects. To determine PEM, we computed the percentage of treatment-phase data above the baseline/Tier-1 median level. Ma (2006) recommended interpreting PEM scores according to the Scruggs, Mastropieri, Cook, and Escobar's (1986) interpretative criteria of effect sizes: A PEM score over 90% was considered highly effective, a PEM score between 70% and 90% suggested moderate effectiveness, and a PEM score slightly below 70% was considered marginally effective.

All three sections had higher average Homework-Accuracy Scores in units during and following treatment than in baseline phases, with improvement ranging from a gain of 8 percentage points (Section 3) to a gain of 22 percentage points (Section 1). PEM scores across sections were 93% for Section 1, 100% for Section 2, and 63% for Section 3 (see Figure 1). Individual PEM scores indicated that the interventions were highly effective for four of the seven participants who completed the study, moderately effective for two of the individuals, and not effective for one individual (see Table 1).

Table 1 provides homework accuracy for individual participants within each section for baseline and treatment/post-treatment phases. The Homework-Accuracy Score mean across participants increased 16% percentage points between baseline phases and treatment/post-treatment phases. The two Section 3 participants, Becky and Stephanie, manifested different trends in homework accuracy: Becky made little improvement in homework accuracy following Tier-2 treatment, whereas Stephanie's homework average increased 13 percentage points after the Tier-2 intervention was implemented. Stephanie's change in performance was similar to percentage gains in PEM participants in Sections 1 and 2.

Table 1 also shows that four of the seven participants who completed the study made gains in homework accuracy during Tier 2 and also met the exam criterion to go to the maintenance phase, and another participant progressed to the maintenance phase after the Tier-3 unit. Analysis of homework scores shows that six of the seven participants who continued their participation in the study met the criteria for homework and exam

Table 1
Homework Accuracy Means, Standard Deviations (in parentheses) for RTI Tiers, and Percent of Treatment Data Points Exceeding the Median (PEM) of Baseline/Tier 1 Data Points

Individual	*Baseline/ Tier 1*	*Tier 2*	*Tier 3*	*Maintenance*	*Treatment/ Combined*	*PEM*
Edith	15 (10.89)	18 (22.77)	51 (23.10)	35 (7.44)	31 (23.37)	73%
Kristin	15 (20.74)	36 (12.87)	N/A	44 (19.98)	36 (18.41)	100%
Sally	7 (8.23)	32 (23.11)	44 (10.77)	N/A	35 (17.30)	93%
Section 1	*12 (13.39)*	*29 (19.58)*	*48 (16.94)*	*40 (13.71)*	*34 (19.70)*	*93%*
Emma	27 (18.65)	36 (2.50)	N/A	44 (14.33)	42 (12.14)	100%
Elisabeth	26 (20.37)	37 (3.30)	N/A	50 (13.56)	46 (12.81)	100%
Section 2	*27 (19.46)*	*37 (2.90)*	*N/A*	*47 (13.95)*	*44 (12.48)*	*100%*
Becky	11 (12.32)	7 (11.98)	21 (6.85)	N/A	14 (11.56)	38%
Stephanie	19 (18.50)	39 (21.65)	N/A	26 (12.52)	32 (17.84)	75%
Section 3	*15 (15.41)*	*23 (16.82)*	*21 (6.85)*	*26 (12.52)*	*23 (14.70)*	*63%*
Combined	**18 (16.09)**	**30 (13.10)**	**35 (11.90)**	**38 (13.39)**	**34 (16.20)**	**85%**

improvement under Tier 2. Two of these participants, Edith and Sally, initially made substantial improvements under Tier 2 but did not sustain this improvement in the subsequent unit, necessitating application of the Tier-3 intervention. The three participants (Edith, Sally, and Becky) who received the Tier-3 intervention made immediate gains in homework accuracy.

Exam Improvement

Exam grades across the seven participants averaged 66% in baseline units and increased to 74% in treatment units. Of the seven participants, four students improved their exam scores from D's in Tier 1 to C's or B's following Tier 2, whereas individuals requiring Tier-3 services made no improvement on exams following baseline. Table 2 shows that students successful in homework and exam improvement in Tier 2 achieved exam scores at or above the class average on treatment/post-treatment exams. Across all participants, deviation from class exam means was -10 percentage points in baseline and -3 percentage points during treatment.

Another important comparison was between participants and eligible students who elected not to participate. Both groups had similar exam averages (low D) in the baseline unit. The average test grade for Units B through E was 72% for participants and 67% for eligible non-participators. Participants averaged a C on all tests except Unit B (high D), whereas non-participants had a grade average of F on Unit B, a grade average of D on Units D and E, and a grade average of C only in Unit C. Ten out of the 31 students who elected not to participate earned an average grade of C or higher on exams B through E ($M = 74\%$), while 21 of those students earned an average grade of D or lower ($M = 64\%$) on these exams. In contrast, 6 out of the 10 participants earned a C or higher across Units B through E ($M = 78\%$), and 4 earned a D or lower ($M = 66\%$).

Discussion and Remaining Research Issues

Overall analyses of the study suggest that an intervention program following selected RTI core principles is feasible at the college level and can lead to improvements in homework accuracy and exam scores. Analyses of results demonstrate that the tier program was effective for each participant in improving homework accuracy, with some participants responding to the instructional adaptations in Tier 2 and the rest showing improvement in Tier 3. Across tiers, four out of the seven participants made gains in homework accuracy following one Tier-2 session, and they

Table 2
Average Exam Percentages and Deviations From Class Mean Across Baseline and Treatment/Post-Treatment Phases for Individuals and Groups

Individual	*Baseline*	*Deviation*	*Treatment/Post-Treatment*	*Deviation*
Edith	66	-12	67	-10
Kristin	66	-12	84	+7
Sally	62	-16	61	-17
Section 1 Group	*65*	*-13*	*71*	*-7*
Emma	67	-3	75	+1
Elizabeth	62	-8	70	-3
Section 2 Group	*65*	*-6*	*73*	*-1*
Becky	73	-8	69	-10
Stephanie	67	-13	87	+8
Section 3 Group	*69*	*-11*	*79*	*0*
Combined	66	*-10*	74	-3

maintained Homework-Accuracy Scores at levels higher than baseline across the remaining units. The participants who received more intensive, one-one-one assistance made higher gains in homework accuracy in Tier 3 than in Tier 2.

A majority of participants made improvement in exam scores between baseline and treatment phases, with an average increase of 8 percentage points. Exam gaps between participant exam scores and class averages were smaller during treatment than baseline phases. Participants performed higher on subsequent exams than non-participants eligible for RTI Tiers 2 and 3. At the individual level, the intervention had somewhat variable effects on exam performance. While individuals who responded to Tier 2 were able to maintain both homework accuracy and exam improvement trends during the semester, those who required the Tier 3 intervention made little exam improvement over their baseline level.

The procedures and results of this study suggest several benefits of implementing selected RTI procedures to enhance the academic performance of college students. Goals for future research in this area are similar to directions suggested for RTI implementation at the secondary level (see Duffy, 2007): Identify interventions that can work across subject areas, find appropriate screening and monitoring tools, consider unique issues of the settings and participants, and provide structural support for collaboration across college professionals. In order to design RTI programs that are flexible enough to fit multiple college settings, researchers should continue to assess the use of various universal screeners, progress-monitoring tools, and interventions adaptable to college courses.

Motivation of College Students

In a previous study, Abrams and Jernigan (1984) found that one of the strongest indicators of first-semester GPA among college students with academic deficiencies was their willingness to seek help from their teachers. Similarly, this study relied on students' willingness to attend tiered sessions out of class. As expected, only a small percent of students identified as needing additional assistance elected to participate in the RTI sessions. Three of the 10 participants who initially participated elected not to proceed to the more intensive intervention tier. The reduced full-participation rate suggests that motivation to engage in the out-of-class intervention was a likely mediating factor driving efforts for improvement on exams. Participants who decided not to proceed with the help sessions made that decision when asked to work more intensely with the researcher.

Universal Screener

Using the first exam as a universal screener has its pros and cons. It served two purposes in that it provided a tool for identifying at-risk students, and it assessed students' understanding of the content in the first unit of the course. The screener was completed within a 50-minute time limit, making identification of eligible students quickly and easily determined. Also, we have found a positive correlation between initial exam scores and performance on future exams for students who took the target course within the previous two years. Specifically, we found that poor initial exam performance (a D or F) was significantly related to performance on each subsequent exam at the $p < .005$ level (Blondin, 2012). Because early performance was related to subsequent exam performance, the first exam may be an accurate and valid screener in courses similar to the one in this study.

Although the universal screener adequately identified students who needed help, we do not yet know the most effective time to include the screener at the college level. Waiting until after the first exam may negatively impact students wanting to get off to a good start in the course. While some students may be motivated to improve subsequent exam scores, others might perceive initial failure as an indication of future failure. By identifying whom to target and offering services prior to the first exam, instructors might enable students to be more comfortable in using effective exam-preparation strategies.

Researchers should continue to explore the use of various universal screeners that efficiently and accurately assess skill levels related to desired outcome measures. One possible screener is performance on critical thinking tests at the beginning of the semester. Compared to others students in the course, the participants in our study were lower in critical thinking (a mean score at the 6th percentile). Because critical thinking consistently has been linked to exam success (Galyon et al., 2013; Williams et al., 2004), researchers should consider critical thinking as a universal screener, starting from the very beginning of the course**.** Whatever universal screener is presumably linked to exam performance, the combination of limited motivation and low critical-thinking scores for initially low exam performers presents a tremendous hurdle for RTI success in improving college students' exam scores.

Progress Monitoring

Research on progress-monitoring tools conducted at the secondary level suggests that more complex skills are harder to evaluate than skills

measured at the elementary level (Duffy, 2007; Ehren, 2008). The PEM analysis showed clear overall differences between Tier 1 and higher tier levels on homework accuracy for Sections 1 and 2. However, the range of participant scores across days in each phase highlights the vulnerability of the study measure to outlier scores. Researchers should continue to evaluate progress-monitoring tools that are shown to reflect target skills in college courses and are sensitive measures of responses to intervention.

In a typical college class, finding a tool that can repeatedly be used accurately and efficiently to measure skill levels can be a challenge. This study used a grading rubric developed by Galyon et al. (2013) to measure homework accuracy. Approximately 2-3 questions were evaluated daily, with a total of nearly 400 questions rated during the study. In large entry-level classrooms, using the current grading rubric to evaluate answers to homework questions might be too labor-intensive to be considered practical by most instructors. Continued research should be conducted on progress monitoring tools that are connected to course exams and are efficient to use. For example, future researchers could explore progress monitoring tools that are already evaluated as part of the curriculum or are efficient to obtain (for example, one question rated per day). Another possibility is using personal response systems, or "clickers," which have the potential continuously to measure students' mastery and progress while engaged in class activities. Clickers have been reported to be an effective formative assessment tool in college classrooms (Judson & Sawada, 2002; MacArthur & Jones, 2008). However, the clicker arrangement would necessitate the potential to separate individual from group responses as well as prevent group responses from affecting individual responses.

Intervention Strategies

Participants who made little progress in this study might have benefitted more from a tier-based intervention following a problem-solving model rather than the standard protocol format used in this study. Unlike the standard protocol model, which uses preselected, evidence-based instructional strategies to develop new skills, the problem-solving model is individually designed to enhance students' current skills and target behavior problems (Fuchs & Fuchs, 2007). In the college setting, a problem-solving model may target specific behavioral obstacles (for example, poor time management, test anxiety) that are interfering with a student's academic performance. Finally, to determine whether effects of various RTI components applied in this study can be generalized across college classrooms, interventions can be designed to target broader test-taking

strategies, time management, and class-preparation skills. For example, Tier-2 services may be broadened to provide general test-taking strategies, and Tier-3 services can then focus on more course-specific variables.

Implications and Conclusions

Analyses of the study's results indicate that interventions featuring RTI principles can be implemented in the college setting if universal screeners are in place, appropriate tools to measure progress are routinely used, instructor time and resources are available, and college students are willing to receive out-of-class services. An important factor in RTI's future applications is whether instructors are willing and equipped to implement such intervention programs in their classrooms. Although universal screeners, progress monitoring measures, and intervention materials were inexpensive to create, we invested considerable time in the creation and implementation of this project. Courses that have resources already available, such as graduate teaching assistants, collaborative professional help, and room availability, might be good prospects for tier-intervention programs. Just as RTI programs at the lower levels rely on resources and assistance from an RTI team to collect data, make empirically based decisions, and implement interventions, programs at the college level may make even greater demands on a team approach. If RTI proves workable and effective at the class-level in a particular college setting, instructional teams may be inclined to apply selected RTI strategies at the program- and even college-wide level.

References

Abrams, H. G., & Jernigan, L. P. (1984). Academic support services and the success of high-risk college students. *American Educational Research Journal, 21,* 261-274.

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23*(4), 362-380.

Blondin, C. A. (2012). *Early identification and improvement of variables related to course success* (Unpublished doctoral dissertation). The University of Tennessee, Knoxville, TN.

Burns, M. K. (2008). Response to intervention at the secondary level. *Principal Leadership, 8,* 12- 15.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic

review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*, 381-394.

Carroll, E., Williams, R. L., & Hautau, B. (2006). Cooperative learning contingencies: Unrelated versus related individual and group contingencies. *Journal of Behavioral Education, 15,* 191-202.

Duffy, H. (2007). *Meeting the needs of significantly struggling learners in high school: A look at approaches to tiered intervention.* Washington, DC: National High School Center, American Institutes of Research.

Ehren, B. J. (2008). Response to intervention in secondary schools: Is it on your radar screen? *RTI Action Network*. Retrieved from http://www.rtinetwork.org/Learn/Why/ar/RadarScreen

Ehren, B. J., & Deshler, D. D. (2010). Using the content literacy continuum as a framework for implementing RTI in secondary schools. *Theory Into Practice, 49,* 315-322.

Foster, L. N., Krohn, K. R., McCleary, D. F., Aspiranti, K. B., Nalls, M. L., Quillivan, C. C., Taylor, C. M., & Williams, R. L. (2009). Increasing low-responding students' participation in class discussion. *Journal of Behavioral Education, 18*, 173-188.

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children, 39,* 14-30.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2010). Rethinking response to intervention at middle and high school. *School Psychology Review, 39*, 22-28.

Galyon, C. E., Blondin, C. A., Forbes, B. E., & Williams, R. L. (2013). Does homework matter? Exam performance in large college classes. *Journal on Excellence in College Teaching, 24*(4), 77-105.

Hautau, B., Turner, H. C., Carroll, E., Jaspers, K., Parker, M. Krohn, K., & Williams, R. L. (2006). Differential daily writing contingencies and performance on major multiple-choice exams. *Journal of Behavioral Education, 15,* 256-273.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.). (2007). Response to intervention at school: The science and practice of assessment and intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 3-9). New York, NY: Springer.

Judson E., & Sawada, D. (2002). Learning from past and present: Electronic response systems in college lecture halls. *Journal of Computers in Mathematics and Science Teaching, 21,* 167-181.

Krohn, K. R., Asiranti, K. B., Foster, L. N., McCleary, D. F., Taylor, C. M., Nalls, M. L., Quillivan, C. C., & Williams, R. L. (2010). Effects of

self-recording and contingent credit on balancing participation across students. *Journal of Behavioral Education, 19,* 134-155.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*, 598-617.

MacArthur, J. R., & Jones, L. L. (2008). A review of literature reports of clickers applicable to college chemistry classrooms. *Chemistry Education Research and Practice, 9,* 187-195.

Mellard, D. (2005). Responsiveness to intervention model. In S. Lee (Ed.), *Encyclopedia of school psychology* (pp. 466-461). Thousand Oaks, CA: SAGE.

Mellard, D. F., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research & Practice, 25,* 217-225.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 schools. *Learning Disabilities Research & Practice, 24,* 186-195.

Reason, R. D. (2009) Student variables that predict retention: Recent research and new developments. *National Association of Student Personnel Administrators Journal, 46,* 482-501.

Reutebuch, C. (2008). Succeed with a response-to-intervention model. *Intervention in School and Clinic, 44*, 126-128.

Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early interventions for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders*, *11*, 260-271.

Shinn, M. R. (2008). RTI at the secondary level. In S. L. Fernley & S. D. Norlin, J. (Eds.), *What do I do when. . . . The answer book on RTI.* Horsham, PA: LRP.

Skinner, C. H., Williams, R. L., & Neddenriep, C. (2004). Using interdependent group-oriented reinforcement to enhance academic performance in general education classrooms. *School Psychology Review, 33,* 383-397.

Tackett, K. K., Roberts, G., Baker, S., & Scammaca, N. (2009). *Implementing response to intervention: Practice and perspectives from five schools. Frequently asked questions***.** Portsmouth, NH: RMS Research Corporation, Center on Instruction.

Turner, H., & Williams, R. L. (2007). Generic vocabulary development and performance on multiple-choice exams. *Journal of College Reading and Learning, 37,* 64-81.

Wallace, M., & Williams, R. L. (2003). Multiple-choice exams: Explanations for student choices. *Teaching of Psychology, 30,* 136-139.

Watson, G. B., & Glaser, E. M. (1994). *Watson-Glaser critical thinking appraisal Form S manual.* San Antonio, TX: Harcourt Brace.

Williams, R. L., Oliver, R., Allin, J., Winn, B., & Booher, C. (2003). Knowledge and critical thinking as course predictors and outcomes. *Inquiry: Critical Thinking Across the Disciplines, 22,* 57-63.

Williams, R. L., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *Journal of General Education, 53*, 37-58.

Williams, R. L., & Stockdale, S. (2003). High-performing students with low critical thinking skills. *Journal of General Education, 52,* 200-226.

Carolyn A. Blondin *currently works as a licensed psychologist with the Korn Learning, Assessment, and Social Skills (KLASS) Center, located on the campus of the University of Tennessee. In 2013, she received her Ph.D. in School Psychology from the University of Tennessee. Her research interests include early academic interventions at the college level and class-wide behavior management.* **Kyle Voils** *completed his undergraduate majors in both philosophy and psychology at the University of Tennessee. He is now enrolled in the School of Law at Northwestern University.* **Charles E. Galyon** *currently works as a psychologist in a private practice where he specializes in diagnosis and treatment of childhood disorders and behavior problems. His research interests include behavioral and cognitive factors contributing to learning, and assessment and instructional practices that best promote successful outcomes for students.* **Robert L. Williams** *is a Professor in the School Psychology program at the University of Tennessee. He received his Ph.D. in Educational Psychology from Peabody College of Education, now at Vanderbilt University. His current research interests relate to prediction and promotion of success in undergraduate courses in teacher education.*