Upper Elementary and Secondary CLASS Technical Manual

The Upper Elementary and Secondary Classroom Assessment Scoring System® (UE CLASS® and S CLASS®) measures are observational measures designed to assess effective teacher-student interactions in grades 4-6 (Upper Elementary) and grades 7-12 (Secondary). This technical manual has two primary purposes: first, to provide information on the psychometric properties of these instruments, including descriptive statistics, inter-rater reliability, reliability of scores, and validity findings; and second, to discuss how to use the CLASS measure to observe classrooms.

The Upper Elementary CLASS and Secondary CLASS dimensions are based on developmental theory and research suggesting that interactions between students and adults are the primary mechanism of student development and learning (Greenberg, Domitrovich, & Bumbarger, 2001; Hamre & Pianta, 2007; Morrison & Connor, 2002; Rutter & Maughan, 2002; Pianta, 2006). The CLASS dimensions are based on interactions between and among teachers and students in classrooms; scoring for the dimensions is not determined by the presence of materials, the physical environment or safety, or the adoption of a specific curriculum. This distinction between *observed interactions* and physical materials or reported use of a curriculum is important because in most settings, basic materials are available and the curriculum or learning standards are fairly well-established. In the Upper Elementary CLASS and Secondary CLASS measures, the focus moves beyond these basic provisions and centers on what teachers *do* with the materials they have as well as the interactions that teachers have with their students.

Overview of the Upper Elementary and Secondary CLASS Measures

The CLASS measures have been developed over the course of nearly two decades to capture the aspects of effective teaching that are most closely aligned with students' academic and social-emotional outcomes. The CLASS framework is based on the theory that interactions between teachers and students fundamentally drive the learning and development that occur within classrooms; effective teachers actively engage students in learning and create environments that are conducive to student learning (Howes et al., 2008). This premise has been borne out in multiple studies involving thousands of classrooms and tens of thousands of students across age levels, from preschool through secondary school.

The CLASS framework groups teacher-student interactions into three broad domains, **Emotional Support**, **Classroom Organization**, and **Instructional Support**. These domains are subdivided into 11 dimensions that are defined in terms of clear, specific, observable behaviors. One additional dimension, Student Engagement, stands alone and is not incorporated into any of the domains, for 12 dimensions in total. To complete a classroom observation, each dimension is assigned a code on a scale from 1 to 7, with detailed descriptors of low-, mid-, and high-range codes.

Dimensions	Emotional Support	Classroom Organization	Instructional Support
	Positive Climate Teacher Sensitivity Regard for Adolescent Perspectives	Behavior Management Productivity Negative Climate	Instructional Learning Formats Content Understanding Analysis and Inquiry Quality of Feedback Instructional Dialogue
	Student Engagement		

The Upper Elementary and Secondary versions of the CLASS® observation measure are closely aligned, and both grade levels include the same 12 dimensions. The two age levels differ in terms of the specific descriptions of how each dimension is expressed in upper elementary versus secondary classrooms. The indicators, behavioral markers, and examples have been carefully chosen to best represent appropriate, effective interactions at each grade level.

Data Sources for the Technical Manual

This technical manual draws on data from several studies, each briefly described below.

MyTeachingPartner™-Secondary Study (Secondary MTP™)

Primary Investigators: Dr. Robert C. Pianta and Dr. Joseph P. Allen (University of Virginia)
Study summary: MyTeachingPartner-Secondary (Secondary MTP; Allen, et al., 2011) is an innovative professional development approach that uses a collaborative consultation process based on the Secondary CLASS manual and web-based resources to provide ongoing, classroom-focused in-service training. Funding was provided by the W.T. Grant Foundation and the Institute of Education Sciences.
Number of classrooms: 78 teachers in grades 7-12
Location: Virginia
Sampling: Teachers who agreed to participate within randomly selected schools
Time of year: Throughout one academic year (August to June) during six intervals
Observation procedure: Teachers were asked to send in recordings over the course of the year. These recordings were to be of any lesson where the teacher was actively teaching. Teachers were asked to send in 40 minutes of video. The videos were coded using the Secondary CLASS measure by trained graduate students. Data reported here are from the first year of the project and include both intervention and control teachers.

Measures of Effective Teaching Study (MET)

Primary Investigator: Dr. Thomas J. Kane (Harvard University/Gates Foundation)
Study summary: The MET project is a research partnership of academics, teachers, and education organizations committed to investigating better ways to identify and develop effective teaching. Funding was provided by the Bill & Melinda Gates Foundation.
Number of classrooms: 1333 teachers in grades 4-8
Location: Nationwide
Sampling: Teachers who agreed to participate in random assignment to observation
Time of year: Throughout one academic year four to eight times

Observation procedure: Teachers arranged to have themselves recorded over the course of the year. Roughly half of the lessons scored were math lessons and the other half were ELA lessons. Teachers worked with site-based coordinators to record only when they would be teaching content for a period of at least 30 minutes. These video were coded using the Upper Elementary CLASS and Secondary CLASS measures by trained coders at the University of Virginia and Educational Testing Services (ETS).

Understanding Teaching Quality in Algebra Study (UTQ-A)

Primary Investigators: Dr. Courtney Bell (Educational Testing Service) and Drew Gitomer (Rutgers University)
Study summary: The UTQ-A project was conducted on behalf of Prince George's County Public Schools by the ETS, RAND Corporation, and the University of Virginia. The purpose of this research was to study the potential of a classroom observation system as a teaching evaluation tool that can help principals and teacher leaders give actionable and meaningful feedback to teachers, specifically in algebra classrooms. Funding was provided by the W.T. Grant and Spencer Foundations.
Number of classrooms: 82 teachers from middle and high schools
Location: Virginia
Sampling: Teachers who agreed to participate from middle and high schools in Prince George County
Time of year: Four or five times throughout one academic year
Observation procedure: Teachers arranged to have themselves recorded over the course of the year during their math instruction. Teachers were recorded for a period of at least 30 minutes. These videos were coded using the Secondary CLASS® measure by trained coders at ETS.

Descriptive Information

How do classrooms score on the CLASS observation measure?

Table 1 provides means, standard deviations, and ranges for CLASS scores from three studies. Note that these data are not nationally representative, and are intended to provide examples of scores across several studies rather than national averages.

Several findings are of note. Emotional Support dimensions tended to be scored in the mid-range, with the lowest scores in Regard for Adolescent Perspectives, suggestive of classrooms that are moderately positive but provide students with few high-quality opportunities for autonomy. Classroom Organization dimensions were scored in the mid- to high range, indicating that these classrooms generally provided students with instructional content or activities during the observed lessons and had few disruptions or instances of negativity. Instructional Support dimensions were scored in the low to mid-range. This suggests that these classrooms provided, on average, only moderately stimulating instruction, with few high-quality opportunities for students to strengthen their higher-order thinking skills or deepen their understanding through back-and-forth exchanges. Also of note is that, in the MET study sample, secondary classrooms were rated lower on nearly every dimension compared to upper elementary classrooms.

Table 1. Descriptive statistics (*M, SD, Range*) on the CLASS® measure across three studies

		Secondary MTP	MET		UTQ-A
			4th-6th grade	7th-9th grade	
	N	78	882	698	82
Positive Climate	*M*	4.23	4.68	3.94	3.94
	SD	(0.74)	(.61)	(.69)	(0.71)
	Range		2.38 – 6.50	1.83 – 6.00	
Teacher Sensitivity	*M*	4.44	4.26	3.86	4.21
	SD	(0.68)	(.55)	(.64)	(0.66)
	Range		2.33 – 6.50	1.67 – 5.81	
Regard for Adolescent Perspectives	*M*	3.49	3.29	2.81	2.86
	SD	(0.66)	(.60)	(.72)	(0.61)
	Range		1.38 – 5.18	1.17 – 5.25	
Negative Climate (reversed)	*M*	6.55	6.68	6.42	6.50
	SD	(0.44)	(.35)	(.51)	(0.46)
	Range		4.62 – 7.00	3.44 – 7.00	
Behavior Management	*M*	5.27	6.01	5.54	5.40
	SD	(0.74)	(.58)	(.69)	(0.85)
	Range		2.88 – 7.00	2.50 – 6.80	
Productivity	*M*	5.08	5.91	5.46	5.47
	SD	(0.64)	(.46)	(.69)	(0.59)
	Range		3.88 – 7.00	2.50 – 6.80	
Instructional Learning Formats	*M*	4.23	4.36	3.80	3.86
	SD	(0.65)	(.52)	(.67)	(0.54)
	Range		2.50 – 6.50	2.00 – 5.90	
Content Understanding	*M*	3.74	3.97	3.47	3.89
	SD	(0.61)	(.53)	(.67)	(0.58)
	Range		2.08 – 6.50	1.50 – 6.00	
Analysis and Problem Solving	*M*	3.01	2.80	2.32	2.00
	SD	(0.63)	(.50)	(.62)	(0.48)
	Range		1.63 – 4.31	1.00 – 5.35	
Quality of Feedback	*M*	3.86	3.76	3.12	3.45
	SD	(0.75)	(.57)	(.66)	(0.79)
	Range		2.00 – 5.81	1.25 – 5.40	
Instructional Dialogue	*M*		3.51	2.82	
	SD		(.56)	(.67)	
	Range		2.00 – 5.50	1.25 – 5.19	
Student Engagement	*M*	4.55	5.08	4.31	4.48
	SD	(0.75)	(.48)	(.63)	(0.73)
	Range		2.94 – 6.50	1.88 – 5.95	

Figure 1 shows how CLASS scores in the MET study were distributed across the scale, and demonstrates that the full range of the scale was used. In keeping with the means in Table 1, this figure indicates that Classroom Organization dimensions (Negative Climate (reversed), Behavior

Management, and Productivity) were the most likely to reach the high range, while Instructional Support dimensions (Instructional Learning Formats, Content Understanding, Quality of Feedback, Instructional Dialogue, and Analysis and Problem Solving) were the most likely to be in the low range.

Figure 1. Distribution of CLASS® Scores in the MET Study

This figure shows the percent of CLASS scores in each point range. Each bar represents 100% of the classrooms. For example, for Positive Climate, 1% of classrooms received a 2, 11% received a 3, 41% received a 4, 41% received a 5, and 6% received a 6.

How do the CLASS dimensions relate to each other?

Table 2 provides correlations between CLASS dimensions from the Measures of Effective Teaching study. Nearly all correlations were in the moderate to high range, and correlations were strongest between the dimensions theorized to comprise the three domains.

Table 2. Correlations among CLASS® dimensions from the MET study

	1	2	3	4	5	6	7	8	9	10
1. Positive Climate	-									
2. Teacher Sensitivity	.80	-								
3. Regard for Student Perspectives	.67	.61	-							
4. Negative Climate	-.58	-.53	-.33	-						
5. Behavior Management	.53	.46	.27	-.83	-					
6. Productivity	.52	.47	.29	-.69	.85	-				
7. Instructional Learning Formats	.72	.72	.63	-.50	.54	.61	-			
8. Content Understanding	.57	.62	.52	-.44	.50	.56	.79	-		
9. Analysis and Problem Solving	.55	.58	.72	-.36	.38	.40	.67	.70	-	
10. Quality of Feedback	.71	.74	.64	-.45	.46	.51	.74	.77	.77	-
11. Student Engagement	.71	.68	.62	-.60	.68	.67	.79	.64	.61	.69

Table 3 provides the results of a confirmatory factor analysis conducted to test the three-factor structure of the CLASS observation measure. Results across three studies showed that a three-factor model fit the data well, with all factor loadings at .73 or higher. Cronbach's alphas for each domain, also in Table 3, further indicate that the CLASS domain scores had strong internal consistency.

Table 3. Confirmatory factor analysis results on the CLASS® measure from three studies

Study	Secondary MTP	MET	UTQ-A
Emotional Support		Factor Loading*	
Positive Climate	.83	.85	.91
Teacher Sensitivity	.94	.86	.97
Regard for Student Perspectives	.79	.77	.75
Internal consistency (alpha)	.87	.87	.90
Classroom Organization		Factor Loading*	
Negative Climate	.82	.81	.80
Behavior Management	.95	.96	.96
Productivity	.89	.85	.78
Internal consistency (alpha)	.88	.90	.86
Instructional Support		Factor Loading*	
Instructional Learning Formats	.83	.79	.93
Content Understanding	.81	.76	.90
Analysis and Problem Solving	.79	.77	.73
Quality of Feedback	.91	.84	.85
Internal consistency (alpha)	.88	.92	.91
Indices of fit			
RMSEA	.09	.11	.16
CFI	.92	.91	.91

Table 4 shows correlations between CLASS domains. Correlations were moderate to strong, with the strongest relation between Emotional Support and Instructional Support.

Table 4. Correlations among CLASS domains from the MET study

	1	2	3
1. Emotional Support	-		
2. Classroom Organization	.53	-	
3. Instructional Support	.80	.57	-

Reliability of CLASS Scores

How reliable are CLASS scores?

Evidence suggests that CLASS scores, assigned by trained, certified observers, are highly reliable. The factor analysis and internal consistency estimates for the CLASS domains, presented in Table 3, indicate that the dimensions comprising each domain tap into highly consistent characteristics of classrooms. When measured in the fall and spring, CLASS scores have low to moderate correlations, indicating moderate stability over time (see Table 5). Finally, when two observers code the same cycle,

they consistently assign scores that are within one point on the scale. The tables below provide inter-rater agreement data from four studies. Although observers have an exact match (i.e. assign exactly the same code) about 30% of the time, agreement within one point on the scale ranges from 64% to 98%, indicating moderate to high levels of agreement.

Table 5. Means and stability of Fall and Spring CLASS® scores in Secondary MTP classrooms

	Fall	Spring	Correlations between Fall and Spring
Positive Climate	4.22	4.26	.49
Teachers Sensitivity	4.59	4.43	.48
Regard for Student Perspectives	3.40	3.49	.44
Negative Climate	1.31	1.48	.34
Behavior Management	5.22	5.34	.49
Productivity	5.15	5.06	.47
Instructional Learning Formats	4.27	4.18	.45
Content Understanding	3.64	3.65	.26
Analysis and Problem Solving	3.06	3.04	.26
Quality of Feedback	3.82	3.84	.47

Note: Secondary MTP = Secondary MyTeachingPartner Study

*All correlations above .10 are significant at $p < .001$.

Table 6. Double Coding Reliability for Secondary MTP

Dimension	% Exact	% Exact + Adjacent	ICC	Linear Weighted Kappa	Quadratic Weighted Kappa
Positive Climate	34.5	79.3	.43	.31	.44
Negative Climate	66.5	95.1	.51	.41	.55
Teacher Sensitivity	30.2	73.8	.35	.29	.38
Regard for Student Perspectives	33.2	78.6	.39	.24	.36
Behavior Management	41.2	89.2	.44	.38	.51
Productivity	33.3	84.3	.38	.33	.42
Instructional Learning Formats	31.3	80.3	.39	.30	.39
Content Understanding	31.2	73.4	.38	.29	.37
Analysis and Problem Solving	30.7	73.7	.33	.22	.31
Quality of Feedback	27.6	72.5	.42	.28	.38
Student Engagement	35.2	81.2	.41	.33	.39

Note: Secondary MTP = Secondary MyTeachingPartner Study *Key*: ICC = Intraclass Correlation Coefficient

Table 7. Double Coding Reliability for MET study

Dimension	% Exact	% Exact + Adjacent	ICC	Linear Weighted Kappa	Quadratic Weighted Kappa
Positive Climate	28.2	74.6	.42	.28	.38
Negative Climate	67.5	95.1	.49	.36	.47
Teacher Sensitivity	28.7	72.6	.33	.24	.33
Regard for Student Perspectives	26.6	67.9	.38	.21	.31
Behavior Management	45.8	85.6	.41	.31	.48
Productivity	39.2	82.4	.28	.21	.28
Instructional Learning Formats	31.9	78.2	.35	.21	.33
Content Understanding	30.1	75.5	.31	.20	.30
Analysis and Problem Solving	29.6	71.5	.26	.13	.20
Quality of Feedback	28.7	71.9	.37	.24	.35
Instructional Dialogue	29.2	74.8	.37	.23	.36
Student Engagement	32.2	78.4	.27	.19	.29

Note: MET = Measures of Effective Teaching *Key*: ICC = Intraclass Correlation Coefficient

Table 8. Double Coding Reliability for UTQ-A

Dimension	% Exact	% Exact + Adjacent	Linear Weighted Kappa	Quadratic Weighted Kappa
Positive Climate	26.5	69.3	.03	.08
Negative Climate	61.5	97.6	.07	.13
Teacher Sensitivity	25.4	65.0	.06	.09
Regard for Student Perspectives	27.3	70.2	.11	.21
Behavior Management	47.0	92.8	.08	.14
Productivity	43.2	87.2	.04	.09
Instructional Learning Formats	27.6	73.9	.02	.09
Content Understanding	31.4	76.6	.15	.25
Analysis and Problem Solving	36.3	84.3	.08	.17
Quality of Feedback	23.4	64.1	.06	.10
Student Engagement	33.6	82.0	.07	.14

Note: UTQ-A = Understanding Teaching Quality in Algebra Study

How do observers become certified on the CLASS measure?

Observers are considered reliable when they can code classrooms accurately using the CLASS® lens. Users must prove their reliability and obtain a certification on the CLASS measure before conducting observations. To become certified, observers attend a two-day CLASS Observation Training. During this training, observers learn about CLASS domains and dimensions, then watch and code multiple, videotaped lesson segments that have been master-coded by a team of CLASS experts. Over the course of the two days, trainees calibrate their scoring to be in line with the master coders'. After the training, potential users take a reliability test, which involves independently watching and coding an additional five videotaped lesson segments. Criteria for passing the test include coding within one point of master codes on 80% of the codes overall, and demonstrating proficiency in each dimension by coding within one point of master codes on two out of five videos for each dimension. Users have three opportunities to pass the test. The reliability pass rates are above 80% for both age levels.

How do observers stay reliable on the CLASS measure?

To maintain reliability to the CLASS measure and prevent coding drift over time, observers are strongly recommended to engage in periodic calibration and/or double coding. Calibration involves watching and coding master-coded video clips so that observers can compare their scores to master coders' scores. Double coding involves two or more observers coding the same lesson segment (live or on video) and comparing their scores after the fact to ensure that both observers are coding the same way. Calibration, because it involves periodic check-ins with master-coded video, is an important option for maintaining observer reliability over time and preventing drift. For long-term projects that include observations spread out over several months or a year, a combination of calibration and double coding may be warranted.

One year after passing the initial reliability test and every subsequent year, users must take a new reliability test to recertify on the measure. This recertification test is structured the same way as the initial reliability test: observers have up to three opportunities to watch and code five videotaped lesson segments and are required to meet the reliability criteria in order to continue using the CLASS measure.

Validity

Does the CLASS measure constructs of importance in classrooms (face and construct validity)?

The CLASS measure was developed based on an extensive literature review on classroom practices shown to relate to upper elementary and high school students' social and academic development. The dimensions were derived based on a review of constructs assessed in classroom observation instruments used in school research, literature on effective teaching practices, focus groups, and extensive piloting. Throughout this process, numerous experts in classroom quality and teaching effectiveness have agreed that the CLASS measures aspects of the classroom that are of importance in determining student performance, suggesting considerable face and construct validity.

How does the CLASS measure relate to other measures of classroom quality and associated constructs (criterion validity)?

Criterion validity assesses the extent to which a measure is associated empirically with other measures of similar constructs. Table 9 presents the results from analyses examining associations between the CLASS measure and various other measures of classrooms. These data come from the MET study. Framework for Teaching (FFT) is a general pedagogical instrument like the CLASS measure. The Mathematical Quality of Instruction (MQI) and the UTEACH Teacher Observation Protocol (UTOP) are math-focused instruments, while the Protocol for Language Arts Teaching Observations (PLATO) is a language arts instrument. All correlations are quite high, suggesting that the CLASS measure is capturing something similar to other instruments of quality teaching.

Table 9. Correlations between the CLASS® measure and measures of related constructs

Measure	CLASS
1. CLASS	-
2. FFT	.88
3. UTOP	.68
4. MQI	.69
5. PLATO	.86

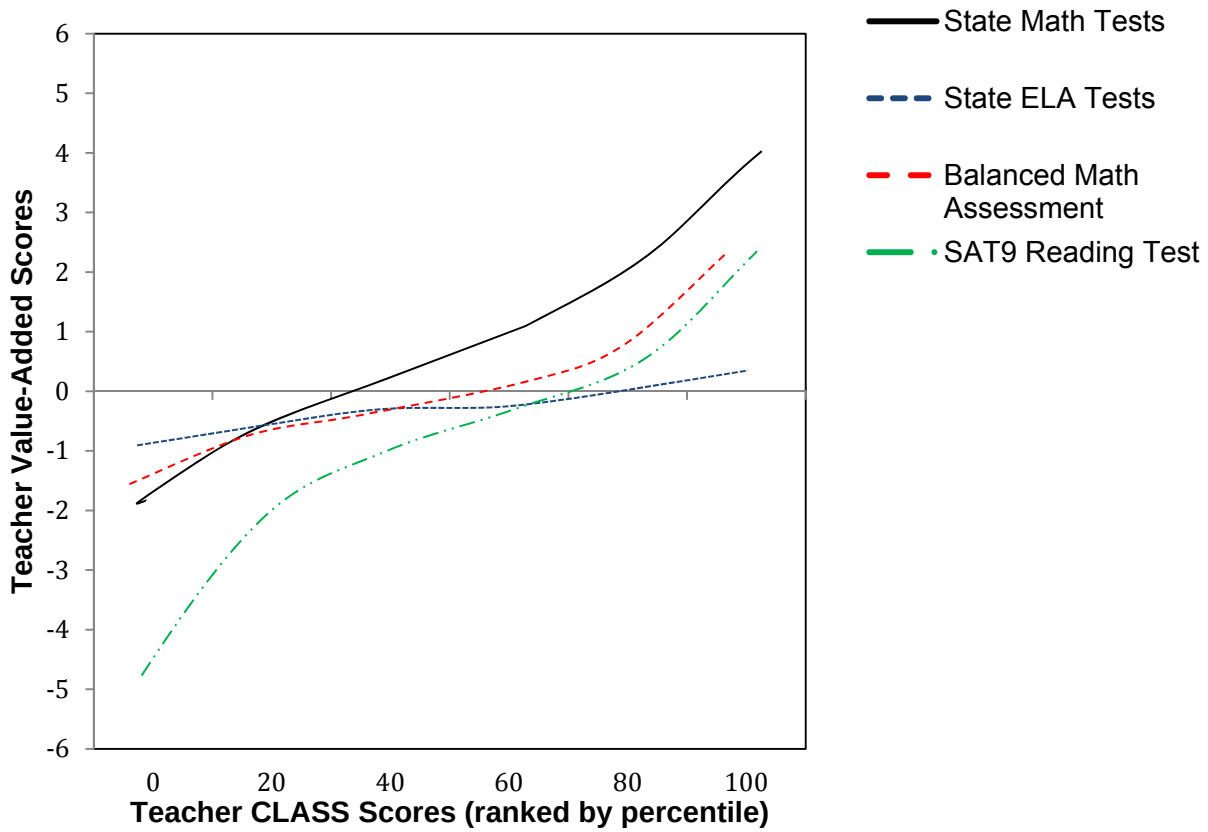
Key: FFT = Framework for Teaching; UTOP = UTEACH Observation Protocol; MQI = Mathematical Quality of Instruction; PLATO = Protocol for Language Arts Teaching Observation.
*All correlations are significant at $p < .001$.

How does the CLASS measure relate to students' academic and social development (predictive validity)?

The CLASS measure was designed to assess classroom-level processes that are directly associated with students' performance. Results from the Secondary MTP study provide evidence that classroom quality, as assessed by the CLASS measure, is associated with gains in children's performance in middle and high school classrooms as assessed by standardized test scores (Allen et al., in press). In this study, the association between the CLASS measure and children's outcomes was assessed after adjusting for a variety of covariates, including grade level, gender, and poverty status.

Data from the MET study provides further evidence for the validity of the CLASS measure. As can be seen in Figure 2, the teachers who demonstrated the types of practices emphasized in the CLASS measure had higher value-added scores than teachers who did not. The figure depicts the average value-added scores (y-axis), expressed in estimated months of schooling gained or lost relative to the average teacher, for teachers at different percentile rankings of CLASS scores (x-axis). The results demonstrate that as teachers' CLASS scores increased, so did the value-added scores.

Figure 2. Estimated Teacher Value-Added Scores by CLASS® Percentile Rank



Summary

Ratings from the CLASS measure provide important descriptive information about the current status of upper elementary and secondary classrooms and also have a predictive value in academic and social outcomes for students. As such, the CLASS measure can be a useful tool for researchers, administrators, and policymakers wanting a standardized measure of classroom processes that are empirically linked to important student outcomes. See the Technical Manual at http://www.classobservation.com for updates.