

May 2023

Adult Rating of Youth Behavior

Measurement Manual



The **Adult Rating of Youth Behavior (ARYB)** asks staff to rate children’s and youth’s *socio-emotional skills* based on behaviors displayed within out-of-school time settings, as observed during several program sessions. ARYB scores are good indicators of how children and youth are likely to perform in settings where they are well supported. Staff should observe each child or youth for at least four program hours before using the ARYB. This manual includes the following sections to support successful implementation and understanding of the ARYB:

Introduction.....	1
Theory of Change.....	2
Adult Rating of Youth Behavior (ARYB)	3
About the ARYB.....	3
How and When to Use	6
Administration and Scoring.....	7
License for Download and Use of QTurn Measures and Manuals	10

Introduction

QTurn's Quality-Outcomes Design and Measure (Q-ODM) toolbox includes a comprehensive suite of evidence-informed measures designed to assess the quality and impact of out-of-school time (OST) programs. This manual provides information about an observational rating instrument designed to assess children's and adolescents' socio-emotional skills. We also provide some technical information about the measure and general guidance about how to administer, score, and interpret it.

The Adult Rating of Youth Behavior (ARYB) was developed from extensive practitioner input and academic research about the socio-emotional behavior of school-aged children and adolescents¹ and the socio-emotional mental skills likely to influence that behavior. It was also designed to help OST program staff (a) think more deeply about how their practices at the point of service influence youth mental skills and the behavioral indicators of those skills and (b) what they can do to learn more about and modify those practices so that they have the intended effects on youth engagement and development.

The ARYB is generally applicable to, and can be used for, any and all types of OST program offerings. The ARYB can also be used with youth classified as having an individualized education program (IEP) or similar designation in the same way you would use it for anyone else. We generally recommend using the ARYB as a primary outcome variable for typical OST program impact studies because we believe that behavioral ratings (as opposed to self-reports) are the most likely to detect socio-emotional skill growth. Please keep in mind that the ARYB does *not* provide a clinical assessment of youth. It is intended only for *lower-stakes* planning, improvement, and program evaluation purposes (e.g., where low scores signal areas of focus and support for youth but not failure, sanctions, or other disciplinary action).

The following sections include:

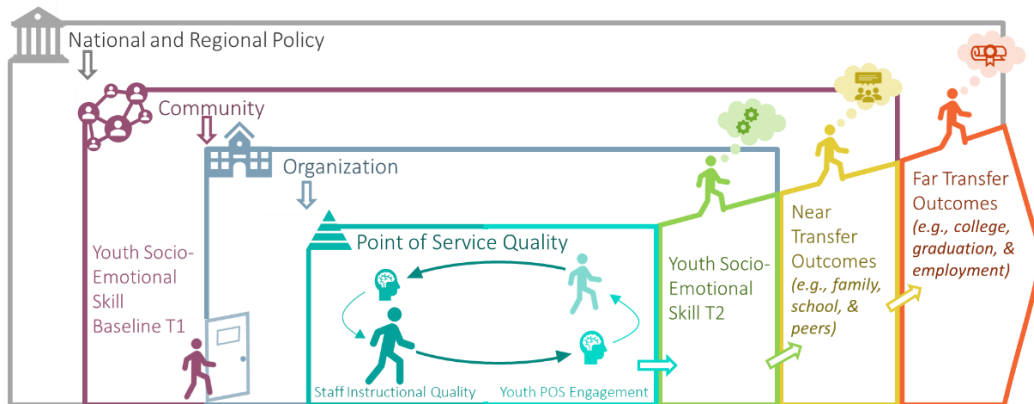
- The theory of change underpinning the measure.
- A technical guide to the measure covering
 - what it is,
 - how it was developed,
 - how and when to use it, and
 - how to administer it.

¹ Hereafter, we use the term *youth* where referring to both school-aged children and adolescents.

Theory of Change

The Adult Rating of Youth Behavior (ARYB) is based on a theory of change (Figure 1) for how youth develop and grow their socio-emotional skills within out-of-school time (OST) settings and then transfer these skills to other areas of their lives (Peck & Smith, 2020). The theory of change is focused on the point of service because that is where adults and youth in OST programs directly interact with each other, so that is where we expect to see the kinds of staff practice and youth engagement that have the most direct impact on the behavioral expression of socio-emotional skills and socio-emotional skill growth. It also includes regional, community, and organizational levels because the quality of staff practice at the point of service depends a lot on the kinds of support staff receive from people working at these 'higher' levels (e.g., professional development opportunities, manageable numbers of youth allotted to each program offering, sufficient resources and staff to help manage the allotted number of youth, etc.). The theory of change also includes near and far transfer elements (e.g., how youth behave in school and the kind of progress they make on their educational or occupational journeys) because that is where we expect to see the ultimate benefits of socio-emotional skill growth.

Figure 1: Theory of Change for Multilevel Cascades of Causes and Effects.



The theory of change indicates that high-quality staff practice at the point of service (e.g., 'meeting youth where they are at') promotes youth mental engagement at the point of service. Youth mental engagement promotes both the optimal behavioral expression of current skills and, through multiple interactions over time (e.g., feedback and mastery experiences), socio-emotional skill growth. In these terms, socio-emotional learning (SEL) is the process through which children and youth develop essential life skills such as empathy, regulating emotions, responsibility, and teamwork. Socio-emotional *skills* are the relatively-enduring mental and behavioral results of these learning processes. These skills play a central role in youth's ability to do well in school, make successful transitions to adulthood, and achieve positive life outcomes, including educational attainment, employment, and good health.

Adult Rating of Youth Behavior (ARYB)

About the ARYB

The Adult Rating of Youth Behavior (ARYB) is an observational rating instrument used to assess *optimal* socio-emotional behavioral skills at the point of service in six domains of socio-emotional functioning (i.e., Emotion Management, Empathy, Initiative, Problem Solving, Responsibility, & Teamwork). Optimal skills are the best someone can do while receiving high-quality support to enact these skills, which is what we expect to find within OST program settings. If your primary goal is to assess socio-emotional skill growth, we generally recommend using observational measures like the ARYB because they focus on optimal behavioral skills, which we view as the most valid and sensitive (to change) information about socio-emotional skills (Smith & Peck, 2020).

The ARYB focuses on specific observable behaviors associated with the socio-emotional skills of youth from 5 to 18 years of age and can be used within the context of any type of program offering in which youth participate on multiple occasions for a consistent purpose. The ARYB was not designed to be an exhaustive measure of socio-emotional skills; rather, it was designed to focus on a few key types of behavioral skill related to each of the following six domains of socio-emotional functioning:

- *Emotion Management* refers to behavior that reflects abilities to be aware of, name, understand, and constructively handle both positive and negative emotions.
- *Empathy* refers to behavior that reflects abilities to feel what other people are feeling and relate to others with acceptance, understanding, and sensitivity to their diverse perspectives and experiences.
- *Initiative* refers to behavior that reflects abilities to take action, sustain motivation, and persevere through challenge toward a specific goal.
- *Problem Solving* refers to behavior that reflects abilities to plan, strategize, and implement sequenced steps.
- *Responsibility* refers to behavior that reflects abilities to reliably meet commitments and fulfill obligations of challenging roles.
- *Teamwork* refers to behavior that reflects abilities to collaborate and coordinate action with others.

In some settings, some of parts of the ARYB may be inapplicable (e.g., teamwork ratings in a tutoring program). We recommend that users select the socio-emotional skill domains that are most applicable to the type and goals of the program offerings in which youth will be rated.

Validity

The ARYB was developed from extensive practitioner input and research associated with the work of Smith et al. (2016a, 2016b) and Larson et al. (e.g., Larson & Angus, 2011; Larson & Brown, 2007; Larson & Hansen, 2005; Larson et al., 2006; Pearce & Larson, 2010; Rusk, Larson, et al., 2013; Salusky, Larson, et al., 2014). Psychometric details, including validity and reliability information, for a previous version of this kind of observational rating system for assessing youth's socio-emotional behavioral skills in the six socio-emotional skill domains used here can be found in the research report by Peck, Smith, Hillaker, Macleod, Roy, Helegda, & Smith (2018), which is available upon request.

Although detailed psychometric information about the validity and reliability of the current version of the ARYB is pending, given the close alignment between the socio-emotional skill domain definitions and the ARYB item content, as well as our experience with previous versions of this tool and the analyses summarized below, the ARYB total and domain scores can be viewed as a reliable and valid measure of youth's optimal socio-emotional skills.

In one sample of youth ($n = 96$, about 76% of whom were age 10-14) who were rated on the ARYB by adult OST practitioners, the alpha coefficient of reliability for the ARYB total score was .96 and, for the six socio-emotional skill domain scale scores, were: .79 for Emotion Management, .76 for Empathy, .91 for Initiative, .90 for Problem Solving, .88 for Responsibility, and .91 for Teamwork. The correlations among these scales ranged from .36 (between Emotion Management and Empathy) to .81 (between Initiative and Problem Solving) and averaged .59. In another sample of youth ($n = 66$, about 80% of whom were age 10-14) who were similarly rated by adult OST practitioners, the alpha coefficient of reliability for the ARYB total score was .94 and, for the six domain scores, were: .78 for Emotion Management, .79 for Empathy, .79 for Initiative, .92 for Problem Solving, .80 for Responsibility, and .85 for Teamwork. The correlations among these scales ranged from .11 (between Emotion Management and Teamwork) to .89 (between Initiative and Problem Solving) and averaged .61.

Across both samples, the discriminant validity indicated by the pattern of correlations is an improvement over previous versions of this tool (e.g., where correlations among scales ranged from .47 to .84 and averaged .65 across 12 samples of youth). The differentiation among the six ARYB scales, along with their relatively high reliabilities, is sufficient to produce both meaningful profiles of youth's integrated set of socio-emotional skills and performance feedback for staff continuous improvement processes (e.g., planning with data). For example, pattern-centered analysis of the ARYB scales for the sample of 96 youth revealed six different socio-emotional skill profiles, with 11% of youth evidencing very strong socio-emotional skill profiles, 8% of youth evidencing very weak socio-emotional skill profiles, and the remainder of youth evidencing one of several distinct moderately-skilled socio-emotional profiles, each signaling areas of staff practice from which youth would most likely benefit.

Analysis

After calculating the ARYB domain scores, this information can be used to inform *continuous quality improvement* (CQI) processes, such as decisions about future training and professional development opportunities or program planning. For example, organizations serving high percentages of youth who score very low in a given domain may decide to focus training and program offerings more closely on staff practices that directly address youth socio-emotional skill development needs in that domain. If the ARYB is also used as a follow-up measure, then the ARYB domain scores can be used to assess socio-emotional skill change. The typical way change is assessed involves comparing the ARYB domain scores from the baseline assessment to the ARYB domain scores from the follow-up assessment. However, it is important to keep in mind that, in addition to what the difference between the baseline and follow-up scores might look like ‘to the naked eye’, the interpretation of change scores depends on a host of additional factors (e.g., the number of youth whose scores were used to calculate the average change score and the amount of variation among youth scores around the average of all the scores), we strongly recommend consulting with a professional statistician or program evaluator before drawing any firm conclusions about the meaning of any apparent differences between baseline and follow-up ARYB scores. QTurn provides such program evaluation services on a contractual basis.

Another way that changes in youth’s socio-emotional skills can be assessed, that complements the typical *variable-centered* strategy of calculating change scores for one domain at a time (as described above), is by using *person-centered* (aka, *pattern-centered*) methods that treat the six domain scores for each youth simultaneously, as integrated skill sets (i.e., their socio-emotional skill *profile*). In this way, youth’s socio-emotional skills at baseline and follow-up are indicated by their socio-emotional skill profiles at each of these time points, and change over time (e.g., growth, stability, or decline) is indicated by how youth move from their baseline profile to their follow-up profile. As part of QTurn’s Q-ODM toolbox, after analyzing the ARYB domain score data using variable-centered methods, we generally provide reports to each organization that show the six ARYB socio-emotional skill domain scores as a set of profiles characterizing the participating youth. For example, we provide reports showing the percent of youth with high-, moderate-, and lower-skill ARYB profiles at baseline and follow-up and the percent of youth who followed socio-emotional skill growth, stability, or decline pathways.

This kind of person-centered information can be used for both CQI and program evaluation purposes, such as assessing the extent to which staff practices are having their intended impact on youth’s socio-emotional skill growth. In addition, where organizations are part of larger networks using the same measure, this kind of socio-emotional skill profile and pathway information can be easily aggregated to the network level providing, for example, the basis for more rigorous evaluations of the impact of program quality on youth’s socio-emotional skill growth. Whether considered at the organization or network level, finding evidence of socio-emotional skill growth using ARYB scores supports conclusions like (a) the youth participating in our program offerings are improving their socio-emotional skills and, potentially (e.g., where program quality data are included), (b) the reason youth are improving their socio-emotional skills is because we are creating high-quality program offerings.

How and When to Use

The ARYB should be used by staff to assess youth's optimal socio-emotional behavioral skills at the point of service, where adults and youth meet regularly for a particular purpose. The observation should take place over a minimum of two sessions and approximately 4 hours of session time. It is useful to observe more than two sessions, if possible, within a two-week window. For example, at the beginning of a program, the ARYB should be used about two weeks (or four hours of program time) from when youth first entered the program in order to assess their *baseline* socio-emotional behavioral skills. This means starting your observations of youth behavior as soon as possible after they first start participating in the program offering. However, if unusual circumstances occur during those first two weeks – such as youth being new to the program and not yet 'settled in' or disruptions to normal program operations due to project set-up challenges that affect the behavior of the youth being observed – it is best to wait for a week or two before starting the baseline observations. In any case, at minimum, the baseline observations and scoring should happen within the first month of youth attending the program offering because their socio-emotional skills are probably already growing during this period and it's helpful to have an accurate assessment of their skills when first entering the program. Baseline socio-emotional behavioral skill information can be used to better familiarize you with the youth attending your program offering, to identify settings where a larger proportion of youth are operating with lower socio-emotional skills, and to plan or modify program offering activities (e.g., to tailor activities to the needs of the youth who are attending).

The ARYB can also be used at both baseline and follow-up, shortly after or near the end of a program period, in order to assess socio-emotional behavioral skill growth. If the purpose of the baseline ARYB ratings is to measure 'pre-test' skill information in order to document socio-emotional skill change with follow-up (or 'post-test') ratings, it is best to collect ARYB information as early as possible (i.e., near the four hours of programming mark) because later assessments may not accurately reflect their baseline skills (i.e., they are likely already growing their skills during the first few weeks of program participation). Although ARYB scores tend to be especially useful for detecting meaningful socio-emotional skill growth, we generally recommend waiting at least three to four months to conduct follow-up ARYB ratings if your primary aim is to assess socio-emotional skill growth. However, in some circumstances (e.g., where program activities are focused explicitly on developing specific socio-emotional skills), follow-up ratings within two months of baseline may yield evidence of socio-emotional skill growth.

Using the ARYB at multiple time points supports understanding of several critical patterns pertaining to equity, such as: (a) higher-skill youth should retain these higher skills across the program period, and (b) lower- or moderate-skill youth should evidence socio-emotional skill growth. Using this approach, information about socio-emotional behavioral skills can produce knowledge, understanding, and insights about both positive skill growth outcomes and the relation of these outcomes to other aspects of programming (e.g., the impact of instructional quality on socio-emotional behavioral skill growth).

Administration and Scoring

ARYB Items

The ARYB includes 24 core items (corresponding to the six socio-emotional skill domains) and two additional “background” questions about how well the adult rater knows the youth they are rating.

The response scale for the core items, 1-24, is:

1	2	3	4	5
Never	Rarely	Sometimes	Often	Always

Background

B1. How long have you known this child or youth?

[1 = < 2 weeks; 2 = 2-7 weeks, 3 = 2-5 months; 4 = 6-12 months; 5 = > a year]

B2. For about how many hours of program time have you observed this child or youth?

[1 = < 4 hours; 2 = 4-7 hours, 3 = 8-15 hours; 4 = 16-32 hours; 5 = > 32 hours or more]

Emotion Management

1. How often did the youth easily manage both positive and negative feelings (e.g., didn’t lash out at others when feeling bad; didn’t brag or gloat when feeling good)?
2. How often did the youth get frustrated easily (e.g., challenging tasks, minor set-backs, disagreements, or critical feedback cause more frustration than expected for a successful program experience)? (R)
3. How often did the youth respond positively when frustrated by challenging tasks, minor set-backs, disagreements, or critical feedback (e.g., didn’t blame others or stop trying; thought about it and tried again or sought help)?
4. How often did the youth display disruptive behavior or withdraw from participation (e.g., talked when they should have been listening; stopped paying attention; stopped participating in an activity)? (R)

Empathy

5. How often did the youth notice when someone was sad, upset, uncomfortable, or feeling rejected or bullied?
6. How often did the youth respond without judgement to other people’s opinions, feelings, or situation?
7. How often did the youth appear to experience the same feelings as another person’s feelings (e.g., not just understand, but feel what someone else was feeling, or empathize with others)?
8. How often did the youth acknowledge or express concern about social biases (e.g., sexism, racism, or any unfair words, actions, or situations)?

Initiative

9. How often did the youth take the initiative (e.g., jumped in to help clean up; volunteered for a challenging task; quick to raise hand in response to a question put to the group; supported struggling peers)?
10. How often did the youth set an ambitious goal?
11. How often did the youth stay focused and on-task despite distractions?
12. How often did the youth push through during a challenging situation?

Problem Solving

13. How often did the youth evaluate alternative plans for reaching a specific goal (e.g., develop alternative plans, or a plan-B, in case things don't work out)?
14. How often did the youth create plans, with multiple steps, for reaching a specific goal?
15. How often did the youth manage their time (e.g., by showing up on time, sticking to a plan, or moving on when a step may not have been completed perfectly but was good enough)?
16. How often did the youth think about, and adjust to feedback about, their past behavior or task performance?

Responsibility

17. How often did the youth finish the task they started, with minimal supervision?
18. How often did the youth do the things they said they are going to do?
19. How often did the youth do the things a trusted adult asked them to do?
20. How often did the youth acknowledge a mistake and take action to address it?

Teamwork

21. How often did the youth help or cooperate with others?
22. How often did the youth encourage others to do their part?
23. How often did the youth seek help from others when stuck?
24. How often did the youth keep track of their own and others' group progress?

Scoring

1. Rater responses to all items marked by "(R)" should be reverse scored, such that:
1 = 5, 2 = 4, 3 = 3, 4 = 2, and 5 = 1.
2. The mean of the response values across all items within a domain should be calculated to produce a single domain score for each youth. If some item responses are missing, scale scores can be calculated as long as there are responses to at least 3 of the 4 items in each domain (but remember to calculate the mean score based only on the number of questions for which youth have actually provided a response). Each of the six domain-specific scale scores should range from 1 to 5.

Instructions

Read the statements about youth behavior and, using the given response scale, mark the circle that best describes how often you saw the behavior *during the past two weeks (or four hours of program activity)*. Please do not skip any items. There are no right or wrong answers.

Protocol

Please read and carefully follow these criteria for observational ratings using the ARYB:

1. Identify the adult who will conduct the ratings of each participating youth's behavior.
 - a. The adult rater should be a person who regularly leads the session in which the youth participates and has had the opportunity to observe the youth for approximately four hours of program-offering time, over a period at least two weeks, before completing the ARYB.
 - b. Although there is no training requirement, we recommend training on the use of the ARYB, and the adult who rates youth on the ARYB should, ideally, be among the best trained in the areas of youth development, social-emotional skills, and adult practices that promote positive youth development and socio-emotional skill growth.
2. Identify the youth who will be rated, and write their ID number on the rating form.
3. Observe youth during program activities for at least two weeks and for *at least* two hours per week of program time. The ratings should be based on direct observations of the youth, considering only behaviors the rater has actually seen. Behaviors that were reported to have occurred in other settings, or were not observed by the rater, should not be scored.
4. With reference to the guidance below, complete ratings for youth using the Ratings template provided.
 - a. Rate only those youth who have been identified (as outlined in step 2 of the protocol).
 - b. Ensure that you correctly enter their ID number.
 - c. Allow about 10 to 15 minutes to complete the ratings for each youth.
 - d. The ratings should be completed during a quiet time where there are no distractions.
5. For most or all of the rating items, we could have added the phrase "when appropriate" because we want to know how often the behavior is observed in situations where the behavior is considered appropriate for that situation.
 - a. For example, for the question, "How often did the youth help or cooperate with others?" using the response scale option "2" (i.e., rarely) does not mean they were working alone most of the time so had few opportunities to help or cooperate but, rather, that of the opportunities you observed, they rarely helped or cooperated.
 - b. Similarly, for the question, "How often did the youth notice when someone was sad, upset, uncomfortable, or feeling rejected or bullied?" using the response scale option "4" (i.e., most of the time) does *not* mean that there have been lots of times when someone was sad, upset, uncomfortable, or feeling rejected or bullied but, rather, that *of the times* something like this occurred, the youth usually noticed those occurrences.
6. There are no right or wrong answers. Please do not skip any items.

License for Download and Use of QTurn Measures and Manuals

This is a limited license covering your use of QTurn LLC's intellectual property (IP) in the form of six measurement manuals (documents) that include the following QTurn measures: *Teacher Practices Instrument (TPI)*, *Adult Rating of Youth Behavior (ARYB)*, *Adult Rating of Youth (ARY)*, *Youth Rating of Socio-emotional Skills (YRSS)*, *Teacher Wellness Instrument (TWI)*, and *Youth Engagement Survey (YES)*. Hereafter, you are the licensee and we, QTurn LLC, are the licensor; i.e., we at QTurn are extending a limited license to use the listed documents in certain ways and not in others. Please read it carefully. By downloading the QTurn measurement manuals, you agree to all terms in this agreement. If you don't agree, don't download and don't use.

Definitions. (a) *User or licensee* – any person or entity that downloads the QTurn measurement manuals. (b) *Copyright* – All of the written, audio, and visual material provided as QTurn content are the copyright material of, and are owned wholly by, QTurn LLC without limitation. You agree not to make copies without permission of the licensor. (c) *Intellectual Property (IP)* is defined as (1) trademarked symbols, (2) all published and/or copyrighted content, and (3) trade secrets, such as private consulting communications and methodology, and all information that could be marked “internal.”

Licensee Rights. (a) This license confers on the licensee the right to use the measures and manuals at a single location for the benefit of your organization. You do not have the right to contract with additional organizations to re-sell or license the content in the measures and manuals in any circumstance, without prior written permission of QTurn LLC. (b) You may not sell the QTurn measurement manuals. (c) You may not create derivative works from the QTurn measurement manuals for use outside of the location and organization for which this license is being granted without the express written consent of QTurn LLC. (d) This end user agreement must be printed with all copies of QTurn IP used at your location and organization.

Fees. (a) This license confers upon the licensee the right to print one copy of this material and to make subsequent copies as necessary to equip other staff at your location and organization. (b) There is no fee. (c) Licensor accepts no other responsibilities or obligations related to the content as to defects during or after download.

Responsibilities. (a) You, the licensee, assume all risk and responsibilities associated with downloaded QTurn measures and manuals and their effective use. (b) You agree to use the QTurn measures and manuals in a way that is appropriate per protocols and guidance. (c) You further agree not to use the QTurn measures and manuals for decision-making that is higher stakes or otherwise poses threat, harm, or risk to anyone at any time. (d) You agree not to compete with QTurn LLC using QTurn measures and manuals and agree not to enter into any kind of agreement with any partner to do the same.

Ownership. You agree that all aspects of QTurn IP including the measures and manuals and all other QTurn materials, audio, visual, methods, techniques, plans for implementation, systems, authorships, and ideas are proprietary IP and/or valuable trade secrets of the licensor and are protected by law. The licensor retains all rights to QTurn IP in copyrights and trademarks, trade secrets, and other intellectual property rights therein. Your possession of QTurn measures and manuals, or any other QTurn IP, does not transfer any rights to you except as set forth in this agreement.

Term. (a) This license begins at download, continues for one year from that date, and will automatically update in one year. (b) If the licensee breaches terms of this agreement, the licensor may terminate this agreement with 10 days written notice.

Limitations. (a) Licensor makes no warranties, express or implied, and excludes any warranty of fitness for a particular purpose or merchantability. No one else can make any warranty or representation about QTurn IP other than as provided in this paragraph. (b) The licensor will never be liable to you or any third party for any damages of any sort nor any form of action (criminal or civil).

Indemnification. You agree to hold the licensor harmless from any claim, damage or cause of action (inclusive of negligence, misrepresentation, error, or omission) or other breaches of this agreement by you.

Your representations. You have done your own research and due diligence regarding QTurn IP and have made your own independent determination of value and risks. No promises or other representation have been made by the licensor or the licensor's agents other than as provided in this agreement. Modifications of this agreement are not valid without the licensor's written consent.

Miscellaneous. (a) The parties to this agreement are independent contractors and nothing in this agreement should be construed to create relations of partnership or joint venture or any other entanglement. You will not make any representations to any third party that you have authority to act in the name of QTurn LLC or to obligate QTurn LLC in any way. (b) If we have a dispute, it will be settled in the courts in Washtenaw County Michigan. (c) You agree that the licensor has made no representation or warranty concerning your success and that the licensor disclaims any warranty or representation as to the success of the QTurn IP under this agreement.