	Topic	Discussion Notes	TPS Reaction	Action	Thurs. Call Topic	ATR	Future TAC Topic	Future TPS Topic	Quarterly Mtg Topic	Board Meeting	Action for CAL/WIDA PM Team	Action for CAL Psychometrics Team	Action for WIDA
A1	Overview	Questions- Lyle – lacks conceptual framework for guiding – comparability of interpretation Haven't articulated warrants about interpretation Maintain scale scores is mechanism – urge to think of claims and warrants – what kind of evidence required to support Gary – may not want to maintain same scale score – want to make sure scale scores mean the same thing Lyle – interpretation – claim is the scale scores is the same Push beyond looking at scores, articulate warrants about interpretation – Lyle – all studies in service of interpretation Gary – represent same construct Jamal- comparability of PB and computer- accommodation- sources of construct irrelevant variance- cannot use as accommodation Assuming 65% of student will be cb, may be higher. Idea is to move to mostly online as soon as possible. Paper is accommodation – working with APIP to build in accommodation Some states not part of PARC, SBAC, but those are all online states Lyle- in report – provide explanation of adaptations exactly Lyle: Need to address certain questions: o What are the AUA claims for ACCESS for ELLs Online? o What kinds of evidence do we collect to support AUA claims? o Articulate about interpretation	~Back and forth between Gary and Tim: Evidence as an outcome vs. evidence as an input. • Steve's reaction: What evidence do we have that the data is showing the same thing? o Need to re-think old assumptions because the Consortium now has 37 states with a more diverse population. Parameters are not stable. Relying on 350 data points may not reflect this. How broad of a picture do we want to paint? o CAL: The representation of the population from the beginning has grown. Can go back and see how the population has differed from the past. ~Anything coming out of the 350 is the initial evaluation. Then look at equating window sample. Shu Jing can send info on the Reading domain Alt ACCESS study. Parameters are robust. • Steve: Consider looking at all domains beyond Reading. Would want this so that the group can feel comfortable saying that the changes in scales is supportable. • Robert: Want a step by step AMAO approach	1. Provide explanation of adaptations for computer based items/tasks in greater detail in report (Lyle's reccommendation) - (Alicia and Carsten) send TIDP and Framework doc to Lyle first. Psychometrics team will summarize the adaptation in the ATR 2. See how ACCESS population has differed from the past now that there are 37 states. (no action - new standard setting. How will new states affect the equating? Shu Jing can send a doc and PPT. WIDA will see how we can work with the SEA Relations department 4. Consider looking at all domains, not just Reading, so TPS group can feel comformtable about saying that changes in scale is supportable. 5. CAL to see how the population has differed from the past now that the Consortium has grown to 37 states. 6. Shu Jing can send info on the Reading domain Alt ACCESS study that shows that parameters are robust to TPS members	~talk about next steps for TPS Quarterly Meeting.		2. Send out one email with follow-up docs and answers to previous meeting before next meeting	~Make sure that TPS suggested agenda topic is something that is helpful to our work.				1. Psychometrics team will summarize the adaptation in the ATR	1. (Alicia and Carsten) send TIDP and Framework doc to Lyle first.
A2.1	Listening - comparability and outliers	Tim- outliers – variation- do we have conceptual model of what outlying data we feel comfortable getting rid of based on mode- and what are effects of removing outliers Last year – TAC recommended we survey students so we did not stack deck – motivated removal Shu Jing – moving to script based – went both ways -look at survey questions- compare student results, - same school, etc. Also, look at different performance, compare to survey- some students perform better, not always worse – not consistent Jamal – to maintain comparability - cb and pb – mode should be the only difference– some other differences beyond just mode (grade clusters changed) – differences which may impact comparability Dorry –specs exist for first grade – look at first grade performance – selected tasks which were appropriate to first grade – no new specs were developed Carol – overall mean performance may change on form with new grade specification Dorry – yes - preparing - looking at difference in performances by grade. Not a norm reference test. Mean performance should fit ability of student – construct has not changed in transition SHu Jing – still measuring same proficiency levels Aki – scale chart – correlation – disatenuated - high approx .8 – with pl, close to 1 thought it was all grade levels combined – quite disperse for a single grade Shu jing – differences in mean – Gary – coefficient of determination not big - comparable charts paper to paper – distributions as part of equating Figure 2 during ft data, one tier- attenuated. figure 3 using operational data – link between paper and computer may not be strong – in order to get more confidence in CB PB, – simulate paper paper correlation and compare to SB-MD change Survey – keep it simple Jamal Are you going to do item level analyses on mode – can do DIF – or Gary – structural model. Is there any way to increase the size – Dorry – can use equating sample Gary Analysis would need to be simple – data reduction – are we measuring the same construct		1. Shu Jing provide technical brief of Media Based vs. SB to TAC (no publicaction date yet). Shu Jing send to WIDA. 2. Look into ways to increase size, perhaps by using equating sample (for Listening). Should we increase the equating sample up from 1000? 3. Look at item invariance 4. address how PB and CB differences impact comparability (like grade cluster change) 5. TAC wants more information on kinds of tasks and differences – specifications in order to get more confidence in CB PB, – simulate paper-paper correlation and compare to SB-MD change. 6. Analysis would need to be simple to determine if we are measuring the same construct across modes – data reduction techniques 7. Also analysis at Item level – structural model across mode – see if there is mode effect 8. Carol – having that information can change test specs – look at test spec across modes									
	Listening - Video	Lyle – applied linguistics – comparability of tasks – to what extent is WIDA exploring what CB allows for testing – offering videos – construct is different from face to face conversation- play around with video – push in that direction – TAC wants more information on kinds of tasks WIDA – video – exploring – currently staying as close as possible to what was done. Moving forward introducing		1. Explore video items									
A2.2	Reading - comparability	Lyle – read ahead – also use common item approach – use operationally? Yes Have been using to link pb test - can use baseline of other linking Gary – common item assume variance across mode is consistent Use both – common person and common item – if common item fits, then move on Dorry – specs haven't changed, used older items Gary – folder effects, passage effects Dorry – research done long ago to show they are independent Carol – scrolling? Dorry one item that scrolled, generally no Gary Empirical question – comparability in folders – something different across mode, how students interact with passage Lyle – item types presented in test and item design plan Gary – in peer review w- transition doc – transition is something guidelines will address Jamal - Analyze including and excluding outliers CAL: If there are different results, then you can say that the outliers may have impact on linking. Then what action should we take? This would not be common person equating. Lyle: Compare to common item Gary – outlier is still there – what is decision rule? Lyle- convince stakeholders – Tim – what is belief about outliers? Any data -set 95% cut – will throw out 15 kids in any analysis. Did not finish – strengthened to have some sort of theory for outliers and apply Carol – but is differential performance construct relevant Dorry – conducted cog labs – addressed in different ways- do something systematic. Gary - need to check comparability in the folders. Is there difference in reading on paper vs. computer screer		1. Gary – common item assume variance across mode is consistent Use both – common person and common item – if common Item fits, then move on 2. Look at comparability in folders – is there something different across mode, how students interact with passage. 3. peer review guidelines – transition is something guidelines will address- peers will make judgment on comparability of transition based on documentation 4. Do analysis including and excluding outliers – compare- how to choose results? Did outliers have impact on linking? What is decision rule? Have some sort of theory for outliers and apply – motivated – differential performance - systemic									

A2.3	Writing - Scoring +/-	Scoring – Lyle – pluses and minuses take on meaning which may not be similar across proficiency levels – scale should reflect intentions built into ratings Point is not to say it is continuous –make scale reflect interpretations – bigger differences between categories in diff prof levels Score report includes descriptors – scale score and interpretation – generic performance level descriptors from standards – generic across 1-12 in standards – plus what does that mean for diff grade levels – Lyle – wants to see rating scale and interpretive scale – are they similar – meaning should be conveyed in final score Carol – hoping for higher levels of agreement than done in past Dorry – data not yet available Tim - % of scores on access and online access – some indication that raters are selecting some ratings more than other (444) – is that similar in the new scale. That could affect how it is equated. Dorry – focus group with raters this week Lyle – what use is being made of analytic rating vs holistic – Gary – research on teacher report – scores by standard – people do not generally use report – structural model – some relationship, but not used by teachers- Dorry – that is motivation for simplifying Lyle – analytic not being used – then why do it? Dorry – simplify- new raters report this rubric is simpler to use Use questionnaire to inform outlier analysis – effect of keyboarding later Goal – same step parameters for all test forms, but investigate Tim – what rules for scores to be used for linking – what is interrater agreement rules for both – Dorry – look at raw ratings – similar ball park – Tim – necap- all samples double scored – what is done here, Dorry – everything double scored for field test – Gary – plusses and minuses on each task – pilot report to see if it has any meaning Score report based on wida members desire for diagnostic information, but no research done on how to use it Lyle – need to say what pluses and minuses are- could be done for different reasons Clear description		1. In report – include interpretive cross walk of score and interpretations 2. In report – include % of scores on access and online access – some indication that raters are selecting some ratings more than other (444) – is that similar in the new scale. That could affect how it is equated. Show how raters are using new rubric 3. Take concern with pluses and minuses to scoring committee – may make raters feel good, but could cause problems with interpretations, Will be presented at next year's TAC 4. Address concern about application across grade and tiers 5. Try partial credit model on each task			3. Take concern with pluses and minuses to scoring committee – may make raters feel good, but could cause problems with interpretations, Will be presented at next year's TAC						
A2.3	Writing - Scoring Reliability	Tim – reliability of scores Dorry a.3.2 – linguistic and qualitative analyses – look at differential for paper vs paper. Tim- go back and double score old scores? That is a lot Dorry – double score those that are out of kilter? Contractor only provide one score, but provide statistics on adjudication Carol - Field test scoring – how many raters- everything double scored- best of best- rasch measurement analysis of double scores – raters with different level of severity		1. Do rasch measurement analysis of double scores – raters with different level of severity									
A2.3	Writing - comparability	Lyle Is keyboarding an issue for raters? Yes plans to do those studies Gary – can identify salient differences between ratings – ran structural models, but is form dependent and tier dependent How do you use it? Lyle – need description of differences – on thin ground Gary how do you know what teachers will do with differences Report for interpretive value- Lyle -Need aua – identify most important claims and provide rebuttal – how to prioritize Tim – control for amount of text? Dorry – there is a maximum online Time Length is not scored – what is language being shown – more should not get higher score- hard to control for – report for comparison with necap- part of scoring rubric – some measures of number of answers before and after In necap - Mode affected scores – kids did better in narrative vs persuasive on keyboard. What are controls for longer prompts on type of writing – Dorry – same specs on ft Gary Composite scale proficiency levels – found was not high enough to look at analyses between content test and access – capture higher end to build scale beyond 5 that is meaningful- to give scores Gary – can find level at which content overtakes language Scale post equating – carol – proposal is to anchor – on both step and task difficulty parameters – can get different results		1. Look at discrimination across spectrum. 2. Jenn – assign her to look at claims for FT 3. conduct analysis on word count of keyboarded vs. handwritten 4. Can we augment FT to make sure we have students at the highest proficiency level?									
A2.4	Speaking - linking	Correlations – affected by ceiling affect – need to look at correlation of R/L together with speaking Lyle Concern about students being routed to too hard a difficulty level Can elicit information at that low a level? Don't want to overwhelm student and they don't know what to do Gary: error at lower level will be greater Carol: It's now a multi-modal speaking test. You've changed the construct of the test. CAL will need to draw out how the delivery of the speaking test is similar and different because the paper speaking test is also in some ways multi-modal with the test administrator and booklet. Treat as a new scale as opposed to restructured scale – equipercentile – but have capped scores to deal with- cannot do this Gary – concern about new test vis a vis AMAO's –will have big affect if we draw out speaking. Currently some states require no domain lower than 5 – this may change – policy folks will need to make this call – result of getting more consistent and more focused sample of speaking. Examine the impact of classification accuracy Lyle- need to be able to justify differencesGary – look at classification accuracy of proficiency and scale score– can do propensity score matching to look at types of students, look at impact of new speaking carsten - scoring methods – locally scored – construct irrelevant variance		1. In report - Need to draw out differences – 4 ways they are different – explicitly show effects of differences 2. Policy implications of new and better speaking scale.									
A2.4	Speaking - reactions to demo	Differences – currently face to face – here more scaffolding, model answer, different construct What is motivation for model answer – type of language and level of detail – may not currently expand and elaborate – Ta listening to answer, not language – cog lab, speak like nina speaks With one on one – teacher can answer questions –and scaffold – can you tell me more Evidence – classifications are very different there must be value added for disruption in amaos More meaningful generalizable performance you are obtaining from performance Validation strategy – what other kind of similar task can be administered to verify construct - expect higher correlation – only one construct – resources to express in each domain – not separate domain test s – hope correlations will go up Lyle - Several different types of language activities to demonstrate their capacity – each domain is not skills, but they are how students express themselves in language. Different methods Jamal - Issue of common core requirement more language – explain elaborate, discuss. Language in general is important not only academic language. Students proficient in general and academic Margo – interaction of multiple domains Domain is primary but is mediated through other domains. How to couch that – composite – one underlying construct – moving toward more authentic ways to capture language proficiency		1. Think about how to move towards more authenti ways to capture language proficiency.									

A3.1	Writing - online and paper	• State policy to decide whether students will take paper or online. Ability to demonstrate best language – know something about student – at end of first operational year – don't want to see differences if groups are different – some qualitative analyses on schools – whether they are different – instead match by proficiency level – student level – need to know why kids are assigned – school versus individual – is there evidence that student was assigned because they may do better. Once all schools offer cb, then only identified students will be accommodated. How to distinguish in linking, equating – identify students who would need accommodation. Cb may not be appropriate for all students. How to identify students with ability that requires paper based Are students performing differentially – there is an iep identifier in dataset – no iep, no accommodation In NH, decision not based on iep – teacher/team decision- parents resist having students identified. Accommodations offered with current paper – know what kids are getting and how will those be carried over to cb. Note – pb is not blanket accommodation Personal needs profile – appropriate accommodations provided seamlessly online Subcommittee looking at accommodations and how we ensure they are equitable Tim: will want to know what kinds of students take paper vs. online. If school decides this for all students rather than case-by-case (accommodations), how does this affect the study results? Dorry: This study can be done after the completion of the 1st operational year. Gary: students who don't have an IEP should not take the paper-based accommodation (so take out from the data set). Tim: students without IEP can still take something with an accommodation. Lyle: Consider what kinds of accommodations are offered for paper version now and what kind of accommodations are offered for paper future and online. • ACTION FOR WIDA: Pass studies done through the AAE SC to the TAC • Carol: Same raters vs. different raters: investigate whether the scoring result is the same • Lyle: keyboarding is construct relevant. • Jamal: add student survey question – Do you have a computer at home? (Too late to add to student survey for 2015 FT) • Lyle and Aki's research idea: Create a dummy variable to look at various groups: those who scored high but did not feel that they're better at keyboarding, those who scored low but feel that they're good at keyboarding, etc. Compare to their handwritten scores		1. Conduct study on students who take paper vs. online after completion of the 1st operational year - differentiate between students assigned based on individual differences vs classroom . (Sandra post-meeting thought: Add research questions about this in end of test student survey for operational test?) 2. Think about how paper version is used whether it's for accommodations or other reasons. Consider what kinds of accommodations are offered for paper now and what kinds are offered for paper in the future. 3. Bring studies done through the AAE SC to the TAC. 4. Conduct additional study of same raters vs. different raters. Are results the same?									
A3.1	Listening and Reading	Two formats – not policy Comparability at scale score level Interpretations – how will we ensure interpretation based on two formats will be comparable Accommodation – should be alternative mode – schools not able to administer online also, not accommodation for all students- affordance rather than accommodation. Change the name of paper form from "Accommodated form" to just "paper form" or something else. Adjustments that do not alter construct – administration - from standards – don't alter the construct Calibrating listen and reading for online separately, then calibrate paper separately? Imposing parameters on paper test – are parameters similar? Note – not the same items Gary Require good sample for field test 350 is a little small Jamal May not be able to do item level analysis- items are different – DIF would not apply Can we include some external criteria – see is structures the same or difference – student performance on entire assessment – have operational access How many observed variable - 18-21 items for factor analysis – different items – claim standards are the same Create testlets based on contents – same test specs – distribution of math same as LASS,,,, Expect loadings on LR PB to mirror CB exploratory or confirmatory- confirmatory to test variance Carol: same student s- within student analysis –2 parallel winsteps – look for comparable difficulty– rule out alternative explanation – means it is mode effect no linking items – common students ensure sample is good- across range of pls – particularly for confirmatory factor analysis there are a lot of items for small number of students could reduce to thematic folder? May not have enough students for confirmatory Number of factors depends on tier Scores dichotomous – do tetracorrelation, use mplus. Alternatives to FA Scores should have same factor structure- same subtraits of domain- should have what we designed – unidimensional Comparability is over different regions		1. Topic should be: how will we ensure interpretation based on two formats will be comparable 2. Be careful on use of word accommodation – support 3. there are a lot of items for small number of students. Could reduce to thematic folder? 4. Comparability with external criteria –try correlation with other domains. Run Two parallel winsteps for cb and pb – should be able to plot PLs- correlation 5. Run confirmatory factor analysis - tetracorrelation, use mplus.							Change the name of paper form from "Accommodated form" to just "paper form" or something else.		
A3.1	Speaking - PB vs. CB	Is mode study worth doing- resources – look at biggest problems – lit review should show there is no issue Creating an operational mode that never existed Half paper based, half cb, one test – counter balanced. responses recorded Either CAL or WIDA score Rater condition study - paper – locally scored will be scores of record. Send responses to be centrally scored. Should not be interaction with administrator Training of local scorers –training materials are being developed – TAs will be trained- scorer rubric – simplified for distributed scoring – pilot study – try it out at pilot – to be used for centralized scorers. Who are raters for locally scored – policy – may change a lot- that's why we made scoring simpler Other types of studies during quality control– take matched students –pick on centrally scored and locally scored –post facto analysis- Raters – should meet standards – but what happens locally is beyond our control Hard to make judgments about spoken language online- hope local raters will use model answer. Note – model answer may be at too high a level – students try to reach same level Lyle - Maybe remove model response down the road – may be part of tutorial– students like model in cob lab – not intimidating. Do they know what model is all about?Extrapolation – generalizability – artificial context – as test becomes better known – take that part out – may be interfering Other TAC members like the model response, and think it replaces the scaffolding of TAs. No followup or rewording of prompt currently – model takes the place of this option Speaking happens in context of conversation – exploring more interactive things – students talk to Nina Responses are edited so raters only hear responses – will raters know which questions – yes – will have all resources they need.		1. How is decision made for comparability – do we need to have all aspects of comparability – will be same task? Yes- what happens if one is not met? 2. Interpretation of performances is highest level- that claim – people will -have to weigh whether issue affect interpretations 3. Look at overall Ratings and scores Concurrent operational – if we know interaction for current than can we demonstrate they are more comparable for new.- does rater task interact with mode – effects of interaction – is it better on new. 4. Conduct lit review to show different mode does not present problem 5. Conduct rater condition study - locally score reponse and send same response to be centrally scored 6. Post facto analysis? 7. Study on model response? (potential student survey question?)									

A3.2	**Keyboarding**	Extrapolation of generalizability rather than meaningfulness – is construct to generalize to domain in which they are writing We have new comers – sife – need to give students opportunity to build kb ability Policy – how are we defining this contruct? Outcome – difference in score kb vs hw Is hw gold standard_ no what is to be compared against – what is changing May make a big difference for individual students Reliability of differences will be very low –what does small difference mean vs large difference – how to interpret Regression probably won't tell much – more latent class problem - Could students take longer online? Summary of survey for ecsc broken by grade/form Number of classes required degrees of freedom – only 350 variables Do two separate analyses – self identified – correlation with computer delivered vs correlation of hw Guidance to teachers of what to consider – brand new students 6 variables – how many factors – insteadlook at very low vs very high – match as dummy variables – two groups of students – look at mean criterion score on test – may be variance in demographics Is this worthwhile? Yes – but keep it simple Propose to wida- replicating scoring study Keyboarding is future of assessment Writing score is most significant predictor of content scores		1. Is KB relationship construct irrelevant? Policy – how are we defining this contruct? 2. The keyboard indicator is only a 3. will student level variables make an impact 4. Regression probably won't tell much – more latent class problem - use difference scores as predictor rather than outcome Is developing indicator of keyboard readinesss worthwhile? The TAC says Yes – but keep it simple 5. Do two separate analyses – self identified – correlation with computer delivered vs correlation of hw 6. Propose to WIDA- replicating scoring study- TAC supports this idea- could same student enter in data? May be sufficient to just type students' responses. – cite studies that have done this in different areas – juxtapose different findings									
A4	**Standards Setting**	Two related but different purposes for standard settings First standard setting – determine a score that represents an English language proficiency criterion. A floor level below which a state may not go States may want to go above it – meeting common definition is level below which no state will go. Can you have multiple points for different grades – analysis between content performance and language proficiency- will have to be smooth – set all grades to same point – everything is scaled to fit that level, so bar will be set at same level for all grades What do we use as impact data for bookmarking – if scale is maintained – use historical access data- can go beyond just using field test data – put cuts by grade level- there are different methodologies – multiple parallel panels? External agency – rfp to reach job. Give them parameters – what are critical issues to be thought about – what evidence will you be showing – Previous standard settings (Tim) problems – went with middle of road – not as connected with proficiency as with political agenda. Should higher eds be on panel for 6th grade – will have lower expectations Results can be skewed by panelists – what does it mean to be ELL in class without support – need to be extremely well defined. Do we make recommendations about appropriate methodologies – Jamal – standard methods - bookmark, body of work, mapit based on set of arguments -how independent should CAL be? Very independent for first standard setting, and CAL very involved for second standard setting	~Include state perceptions of policies that each of the 35 states have adopted. Get anecdotal feedback and provide as a background discourse document. This Qualitative data can help support empirical information to smooth over each grade level. ~Panelists would then feel that they are being heard and considered. ~Share with an articulation panel ~Any scale scores that is re-worked should be able to put in terms of the current. Consider in the 2nd Standards Setting study. Maybe the 2nd Standards Setting study is the articulation panel. Qualifications for the 2nd study is more stringent because they need to have had student experience.	1. WIDA will contract a 3rd party to propose a study by July 2015 (1. Single cut score and performance standard for reclassification eligibility. 2. Multiple cut scores to interpret scale scores per WIDA ELD Standards) Check with Gary if Research has done anything on this. WIDA will need to issue an RFP announcement soon to do this! 2. Definition of common cut score - A score that represents an English language proficiency criterion 3. what does it mean to be ELL in class without support – need to be extremely well defined? Need to be really careful- potentially very political and very sensitive 4. what do we know about scores, and how successful they are – other measures – feed into first standard setting – what would that be? 5. (TPS) Include state perceptions of policies that each of the 35 states have adopted. Get anecdotal feedback and provide as a background discourse document. This Qualitative data can help support empirical information to smooth over each grade									
B1	**Drift**	o Lyle's food for thought questions: How many of "those" do you need to address? Is ACCESS leading reclassification decisions? How does this affect comparability? How many stakeholders would be concerned about this? Ask for their feedback. Have we done the best we can? o Carol: underwhelmed by MT's report. It was trying to create best case scenario rather than year to year overtime analysis. 2/3 of rerated scores show increases – results do not seem to support conclusion that there is no drift		1. Get information from states about whether it impacts students – are more students exiting?									
B2	**ATR**	Carol- Much better – clearer, easier to read Lyle – want focus to turn to full AUA • For ACCESS 2.0, we'll revisit the design of the ATR.		1. Get feedback from states and other users (TPS?) 2. Discuss how to turn towards AUA in the ATR 3. Re-design ATR for 2.0		3. Re-design ATR for 2.0							
C1	**Alt ACCESS**	Small differences between equating and population give TAC confidence in small sample in field test TAC is ok with leaving series 102 alone and update series 103. No new items		1. Leave series 102 alone and update 103.									
	Early Years	Once you have this early years language assessment going , would it have potential for ACCESS? –could it be used instead of using listening and reading scores to indicate where to start on ACCESS online speaking test? – For students designating at very beginning – there is a home background, other information beyond screener improve accuracy of screener- home language survey – access is only one piece of information – multifaceted. • Lyle: No need for separate AUAs for each component of the assessment • Lyle: this is an excellent start. Will share comments with Alicia • Lyle's group has developed a software to help create AUA	under what circumstances would you choose which type of the screener? Need to be clear.	1. Need claim 3 for each screener Maybe claim 4. 2. Under what circumstances would you choose which type of the screener? Need to be clear.									
	Other TAC meeting-related suggestions	Margo- Communication – consider condensing and sharing the readaheads with states to demonstrate how thoroughly we have thought through issues and types of analysis – white paper? – one pager per topic? Dorry - Jenn start to build aua for field test	~Advice to WIDA and CAL: Avoid informational discussion. At least present 3 questions per topic to get to the heart of the matter. ~Allow TPS SC members to speak at TAC meeting ~Should DRC participate? What is their role in the TAC and TPS? ~Request DRC to provide weekly reports showing evidence of quality control	1. Margo- Communication – consider condensing and sharing the readaheads with states to demonstrate how thoroughly we have thought through issues and types of analysis – white paper? – one pager per topic? 2. Jenn start to build AUA for Field Test 3. (TPS) Next TAC meeting, avoid informational discussion and allow TPS members to speak. 4. Determine whether DRC should participate and to what degree? 5. (TPS) Provide list of reports that WIDA is already requesting of DRC and provide to TAC and TPS for their sign-off (reports show evidence of QC)									
TPS	**TPS Purpose and Responsibilities**	• Purpose: Get more state input in technical decisions • Responsibilities: ~Robert Lee: Need to get heads together about measuring assessments. Work with Gary Cook on this ~Report back to the EC. Simplify, translate, and amplify to the EC. When we propose ideas to EC, it is after consultation with TPS members ~Pass information from respective states to WIDA and CAL ~Provide feedback on read-ahead materials		1. Write Mission Statement and send to members 2. Vendor QC - consider inviting a SC member 3. CAL QC review - consider inviting a SC member									

	TPS Structure	~Ideal number of TPS members is 6-9 representatives ~want to have TPS be able to speak up at TAC ~How to report out to the EC? ~One person in this group will report to the EC in addition to some kind of read-ahead. ~Steve proposes a rotating representative ~Can join by phone or in person ~Half an hour report • Name of the SC ~Psychometric and Policy Advisory SC (PPASC) ~Psychometric Advisory SC (PASC) ~Assessment Design and Development Advisory SC (ADDA) ~Measurement Advisory SC (MASC) ~Assessment Policy Advisory SC (APASC) ~Validation Advisory SC • Timeline of Meetings: ~ Have Quarterly Meetings (counting the 2 TAC meetings) ~other 2 meetings are in the form of webinars (Two 1.5 hr webinars) to discuss issues (1st 3 bullets of the responsibilities slide) ~Subcommittee help come up with some agenda topics ~Standing Topic: Reviewing Equating Procedures ~Late Feb 2015 potentially ~Face-to-Face in the Fall TAC (Nov. or Dec.)		1. Ask SC to vote on name of the group when we send out the Mission Statement 2. Send Doodle poll to set date for next meeting (late Feb. 2015 potentially) and provide info on next TAC dates. 3. Invite a TPS member to attend EC via call?									
	Next TPS Meeting Agenda	~1st Webinar discuss objectives for the year. ~Think about what info you would like from the new Test Development vendor (Dave Chayer) ~Standing Topic: Reviewing Equating Procedures ~Research on growth model (Gary's work) and providing feedback (priority as expressed by Steve and Robert) ~Take in front of the TAC and the SC measures of growth in order to present full recommendation to the EC. Urge Gary to pilot student growth percentiles in 2015. Consortium-level growth percentiles. ~QC – explore 3rd party QC procedure ~MI has AES conduct the QC. Examine if the delta is big enough ~WIDA would like to issue an RFP. The SC can potentially help! ~Who is the arbiter? This committee. ~Quality control plans for automated scoring and hand-scoring. Help solidify the language from the DRC RFP. ~With embedded field test, we will be able to use data from the field test to do pre-equating. Discuss what near future equating will look like ~Do we want a larger equating sample? How does this work with the reporting window? ~See what criteria is needed to determine sample ~MA and MI both have done this:Consider post-equating and then used that as parameters for pre-equating. Write up and submit to the TAC. This can mitigate drift. Field Test data is set against the post-equating data rather than pre-equating data. ~Acknowledging a better estimate ~Work on a proposal together in the next SC Quarterly Meeting call. ~Review the blueprint design of the embedded Field Test (test map, refreshment schedule, test specifications) – Also read the Test and Item Design Plan. ~Randomly selected students receive embedded field test item ~Innovative item types: hot spots, drag and drop, etc. What do these mean for placement? ~Alternate ACCESS – will not be on the same scale. Discuss Proficiency Level interpretation. Work on new version of Alternate ACCESS.		1. Shu Jing: CAL can send PPT presentation from 3 years ago that explains the process of equating. Keep equating sample as stable as possible. 2. Consider post-equating and then used that as parameters for pre-equating. Write up and submit to the TAC. This can mitigate drift. Field Test data is set against the post-equating data rather than pre-equating data. 3. Carsten send Steve (rest of SC) document about current Proficiency Level.									