Measuring Implementation Fidelity using Computational Methods

Carlos G. Gallo, Ph.D.

Presented at:

Institute of Medicine:

Washington D.C., November 5-6, 2014

NORTHWESTERN
UNIVERSITY

GOALS

- Challenges in measuring fidelity in prevention practice

- Can computational linguistics help?

- Proof of concepts:

  - Familias Unidas (Transcripts)

  - Good Behavior Game (Audio)

- Lessons learned and next steps

NORTHWESTERN
UNIVERSITY

	Efficacy Trial	Effectiveness Trial	Real World Implementation
Facilitator is a…	research team	school counselor	school counselor
Fidelity monitoring is..	high (by research team)	moderate (by research team)	lowest (by **local agency**)
Affordability is..	**yes**	**yes**	**no**

NORTHWESTERN
UNIVERSITY

CNN Tech

Home | TV & Video | U.S. | World | Politics | Justice | Entertainment | **Tech** | Health | Living

Computer finishes off human opponents on 'Jeopardy!'

By **Jason Hanna**, CNN

February 17, 2011 5:50 a.m. EST | Filed under: **Innovation**

$1,000 $2,000

KEN   WATSON
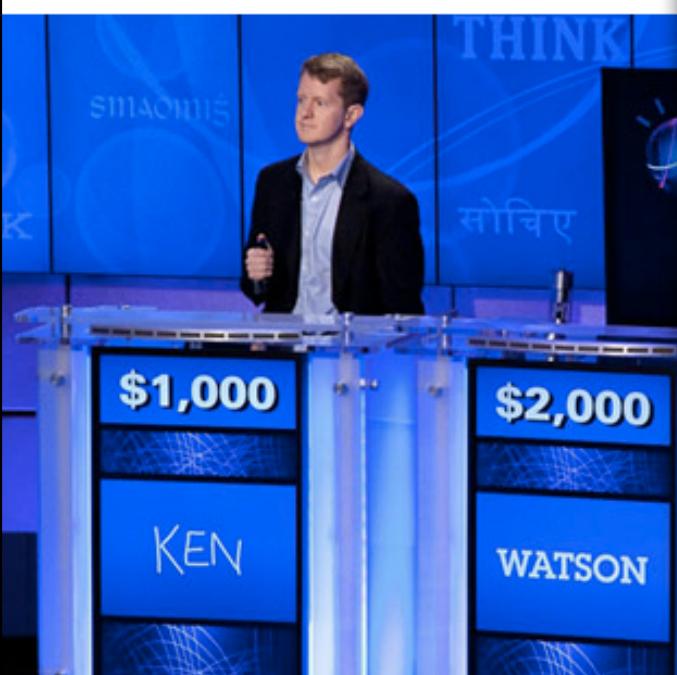
Skype's Translator preview goes live for Windows 8.1 users

by **Billy Steele** | @wmsteele | November 3rd 2014 at 4:39 ...

💬 26

skype

Gurdeep Singh Pall said
Ja. Ich plane derzeit dort i
verschieben, wenn alles ge
yes, i am currently planning to move there somet

We first heard abou
and that it should ar
keeping with that ti
up a preview for Wi

ETS®
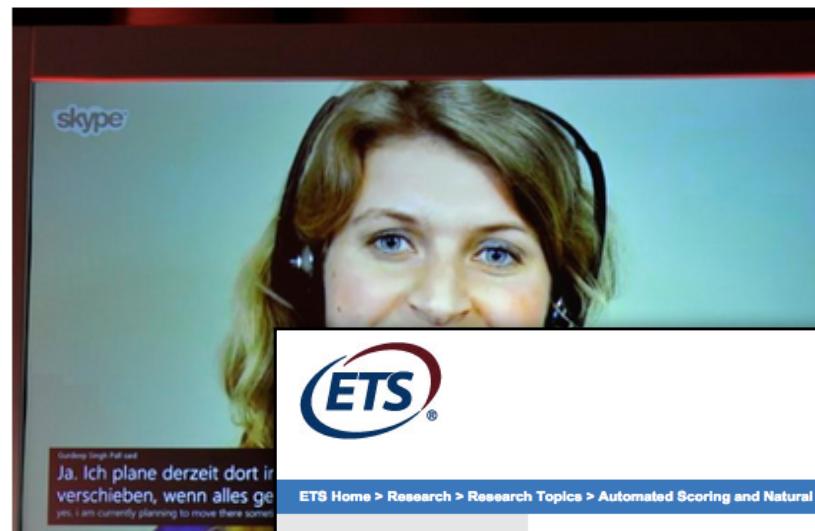
ETS Home > Research > Research Topics > Automated Scoring and Natural Language Processing > Writing Quality

Research Home

Capabilities and Services

Research Topics

Human Capital and Large-scale Assessment

Assessing People with Disabilities

Automated Scoring and Natural Language Processing

Speech

> Writing Quality

Automated Scoring of Writing Quality

The *e-rater*® automated scoring engine is ETS's proven capability for automated scoring of expository, persuasive and summary essays, currently used in multiple high-stakes programs. The *e-rater* engine is used in combination with human raters to score the Writing sections of the *TOEFL*® and *GRE*® tests, as psychometric research has demonstrated that this combination is superior to either machine scoring or human scoring on their own. It is also used as the sole score in lower-stakes contexts, such as formative use in a classroom setting with ETS's *Criterion*® online essay evaluation system. The *e-rater* engine is used to generate the individualized feedback provided to users of *Criterion*. The *e-rater* engine addresses the need for essay scoring that is reliable, valid, fast and flexible, as more and more testing programs, including large-volume state testing, move to online delivery and adopt essay-based tasks for writing assessment.

Human-based method for measuring fidelity

10% sessions

Fidelity trained Human Rater

(weeks or never)

macro-codings expensive

Recorded Intervention

Fidelity Codings
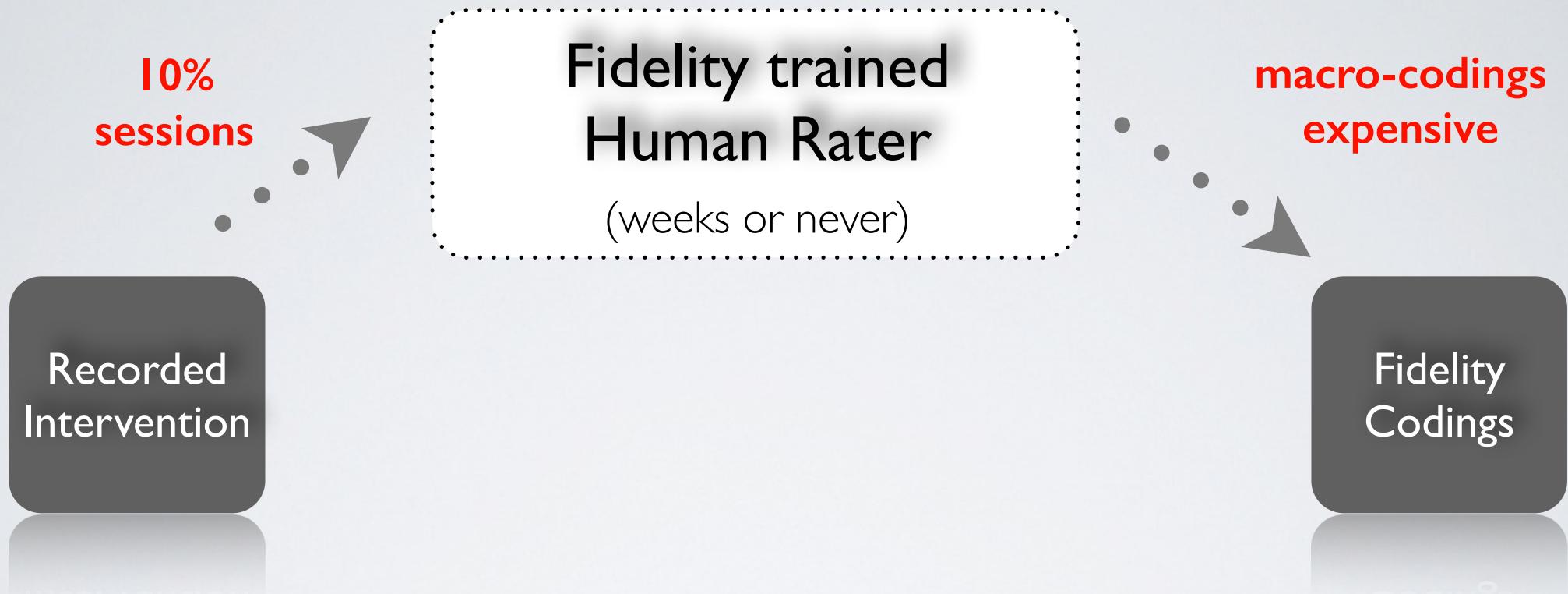
Machine-based method for measuring fidelity

Machine-based method for measuring fidelity
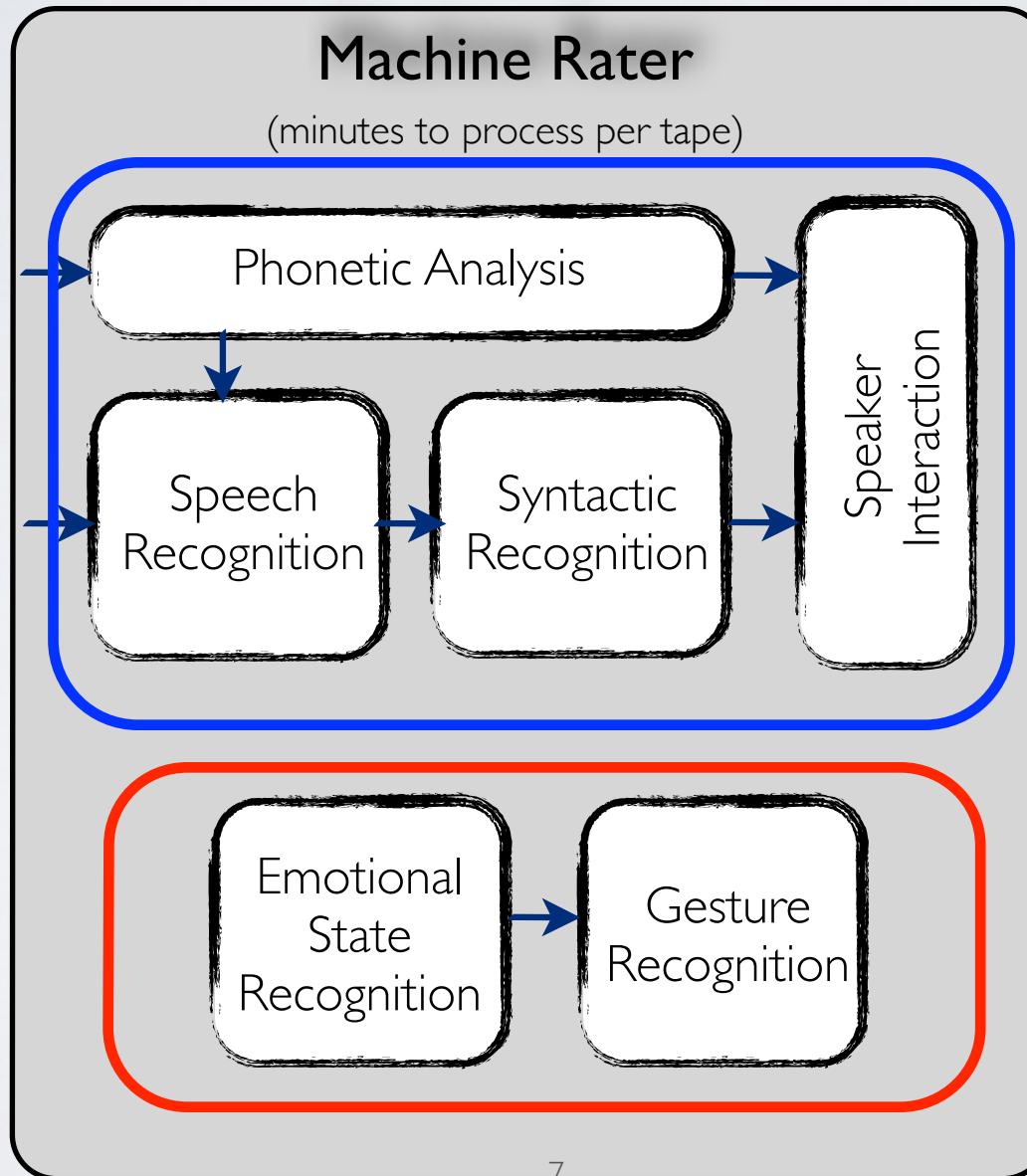
# Monitoring and Feedback System
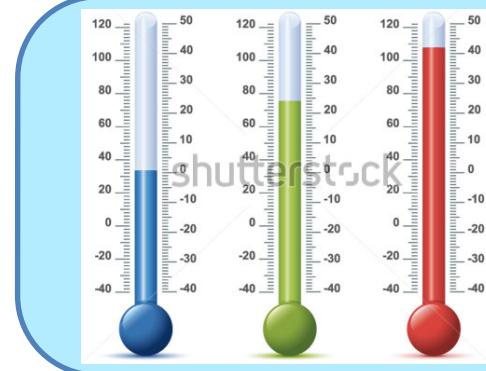
60% Self - (previous)
70% Self - (current)
80% Peers

LINGUISTIC CUES
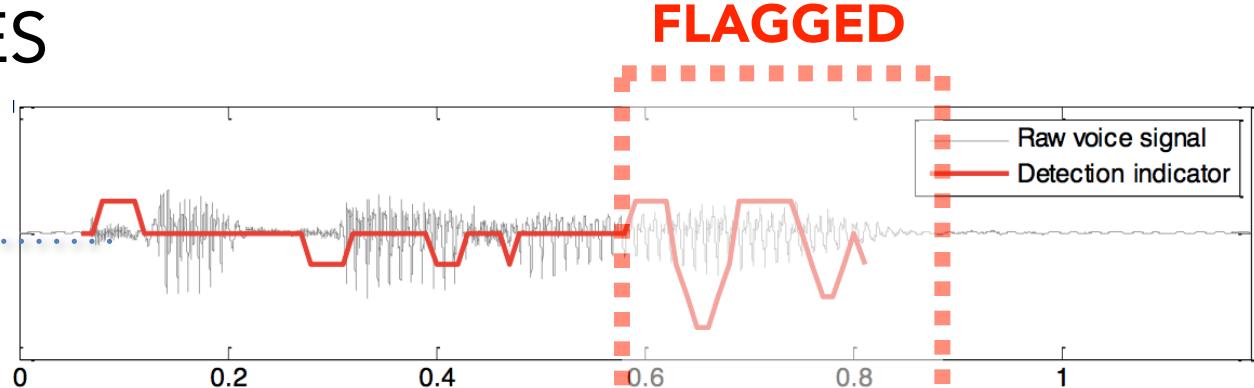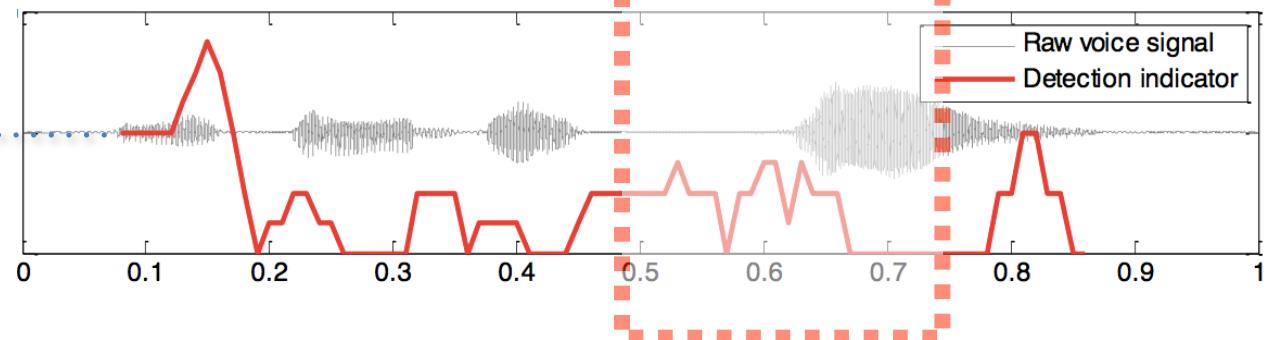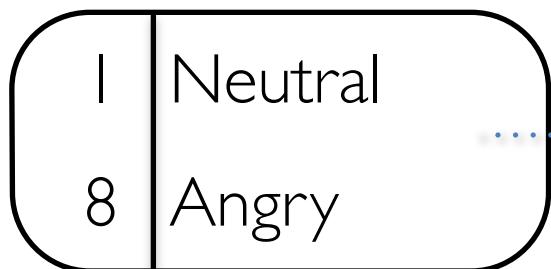
3	High Fidelity
2	Low Fidelity

FLAGGED

Low Fidelity Angry

Raw voice signal
Detection indicator

NON-LING CUES

1	Neutral
8	Angry

Raw voice signal
Detection indicator

A Computational Future for Preventing HIV in Minority Communities: How Advanced Technology Can Improve Implementation of Effective Programs

Brown, C. Hendricks PhD[*]; Mohr, David C. PhD[†]; Gallo, Carlos G. PhD[*]; Mader, Christopher BS[‡];
Palinkas, Lawre[...]
Pantin, Hilda Ph[...]
Mitsunori PhD[‡‡]
PhD[¶]; Villamar[...]

ORIGINAL PAPER

Blending Qualitative and Computational Linguistics Methods for Fidelity Assessment: Experience with the Familias Unidas Preventive Intervention

Carlos Gallo · Hilda Pantin
Guillermo Prado · Maria Ta[...]
Gracelyn Cruden · C. Hend[...]

Gestural cue analysis in automated semantic miscommunication annotation

Masashi Inoue · Mitsunori Ogihara · Ryoko Hanada ·
Nobuhiro Furuyama

CASE STUDY 1:
FAMILIAS UNIDAS (GALLO ET AL. 2014)

- Familias Unidas is an evidence-based parent training intervention for Hispanic youth.

- It is delivered in family visits at home by a school counselor (effectiveness trial).

- A parent and adolescent and school counselor (facilitator).

- Bilingual context.

NORTHWESTERN
UNIVERSITY

MEASURING COMPETENCE IN FAMILIAS UNIDAS

- School counselors *should* **engage/join** family members by asking questions, communicating acceptance, respect, and trust.

- We focus on a dimension of the **competence called** *Joining*.

- Measuring fidelity is expensive and laborious ($800/session) in the effectiveness trial.

Linguistic structures linked to joining

|		Worse
Statements	"You like school" (Subject + Verb + …)	
Yes/No Questions	"Do you like school?" (Aux + Subject + Verb)	
Open-Ended Questions	"Why do you like school?" (Wh + Aux + Subject …)	Better

NORTHWESTERN UNIVERSITY
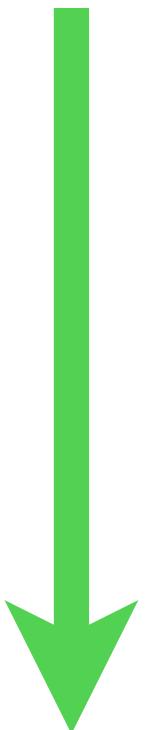
HUMAN AND MACHINE RATER AGREEMENT

Utterance Level:
(Kappa = .83)

Session Level:
(Correlation = .75)
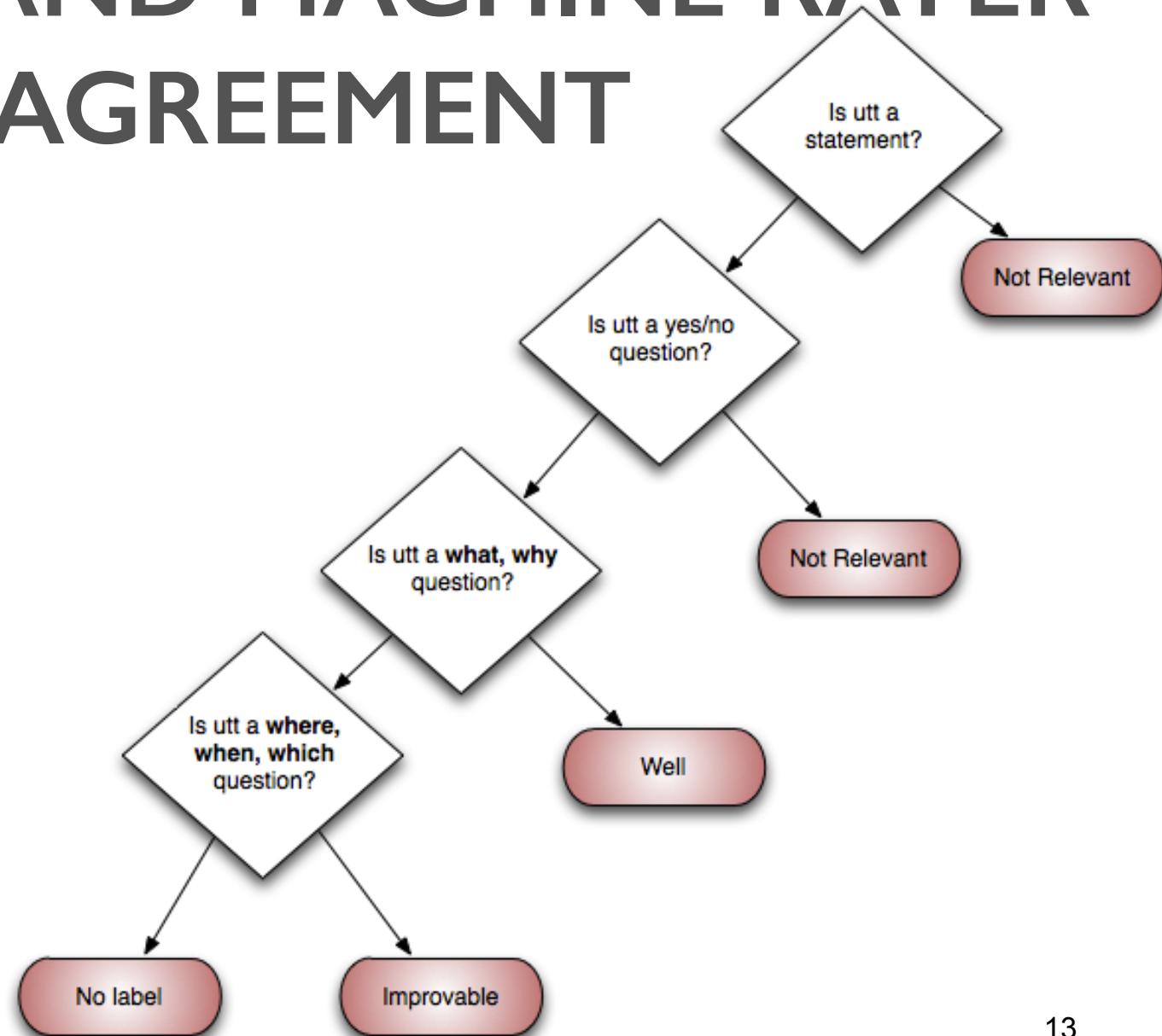
CASE STUDY 2:
GOOD BEHAVIOR GAME (GBG)
(BROWN, MOHR, GALLO ET AL. 2013 JAIDS)

- GBG is an universal classroom behavior management strategy for 1st grade teachers.

- GBG impacts the adolescent and young adult drug abuse, sexual risk behavior, delinquency, and suicidal behavior (Kellam et al 2008, DAD).

NORTHWESTERN
UNIVERSITY

MEASURING COMPETENCE IN GBG

- Teachers *should* provide correctives to children in a neutral tone (not angry, not frustrated, not sad).

- This is a key competence measure in GBG, not collected unless a coach is in the class.

- Coaches find that speaking neutrally is a challenging thing to do.

Guidance from Jeanne Poduska and Gail Chan (AIR)

NORTHWESTERN
UNIVERSITY

MEASURING COMPETENCE IN GBG

We focus on a dimension of competence in speaking to children neutrally.

**Phase 1:** Computationally distinguish emotions in by using audio clips.

NEUTRAL

SPECTROGRAM
"2001"
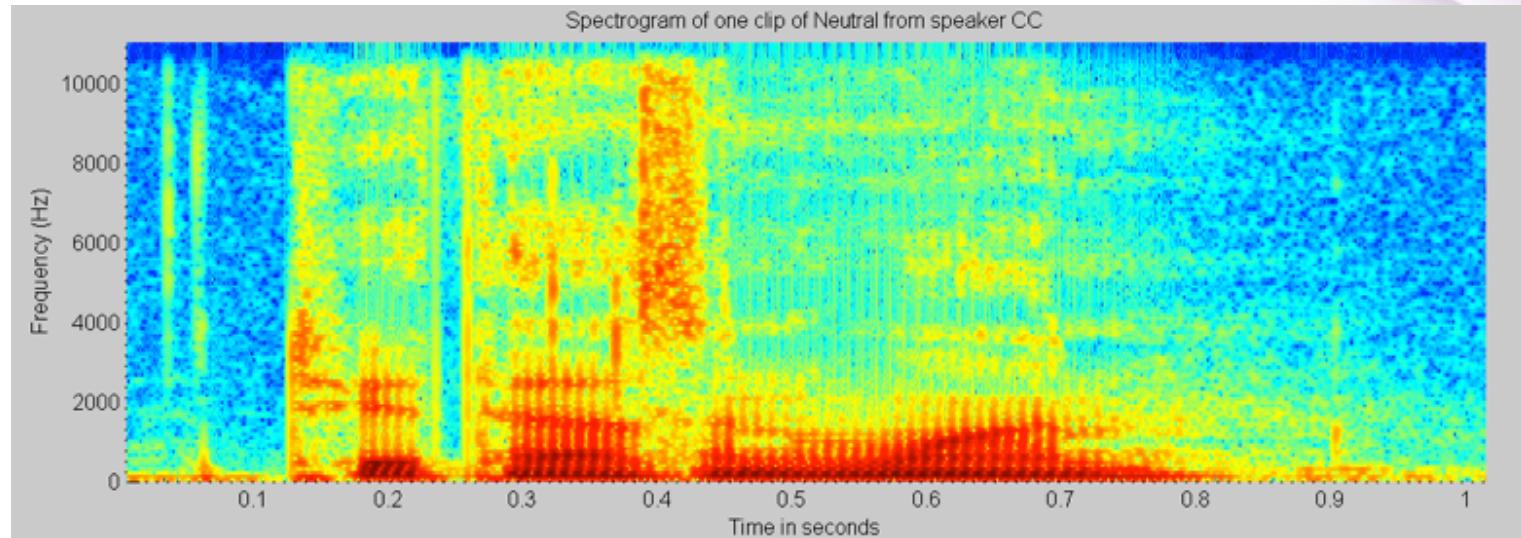
ANGRY

SPECTROGRAM
"NOVEMBER 3RD"

Spectrogram of one clip of Neutral from speaker CC

Spectrogram of one clip of HotAnger from speaker CC

NORTHWESTERN
UNIVERSITY
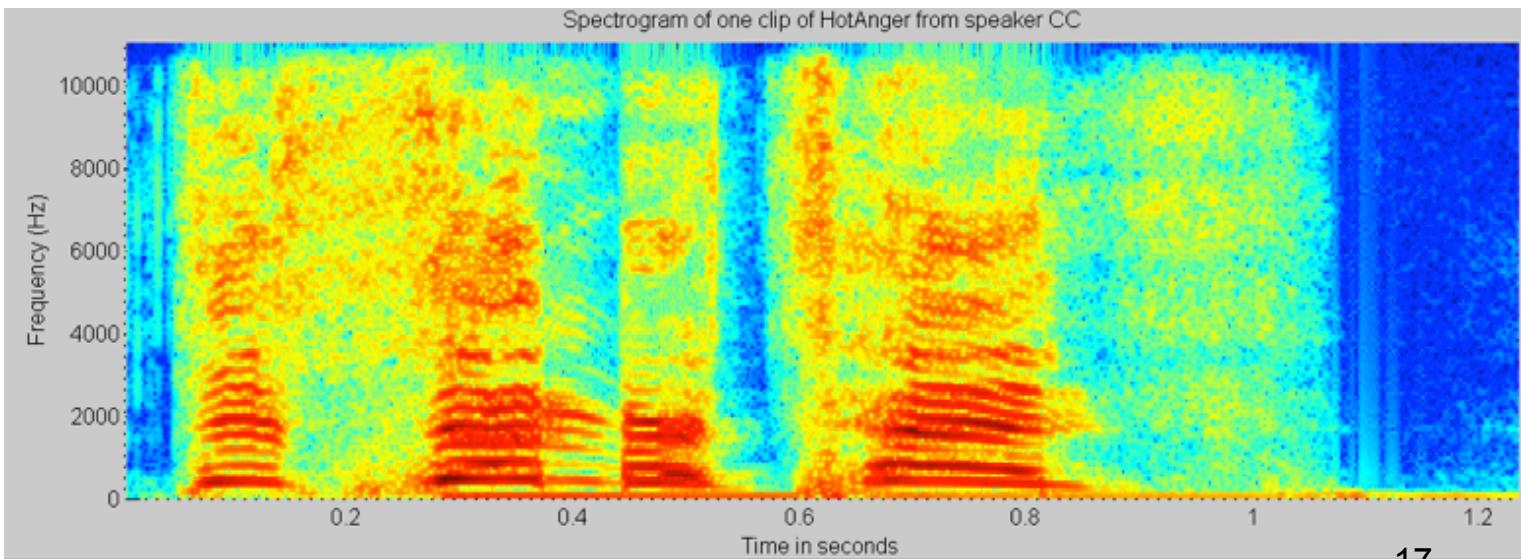
RESULTS

Classification Neutral vs. Anger.   Accuracy = 98%

SPOKEN

	Neutral	Anger	Error
Neutral	78	1	1.2%
Anger	2	138	1.5%

LABELED

Classification Neutral vs. Emotional.

SPOKEN   Accuracy = 87%

	Neutral	Emotional	Error
Neutral	71	8	11.3
Emotional	16	89	18.0

LABELED

NORTHWESTERN
UNIVERSITY

18

SUMMARY

- Technology can be used to assess fidelity cheaply, accurately, and completely.

- Proof of concept: *Familias Unidas* and *GBG.*

- Approach is generalizable to other interventions.

NORTHWESTERN
UNIVERSITY

THANK YOU!!

Carlos G. Gallo

CARLOS.GALLO@NORTHWESTERN.EDU

Department of Psychiatry and

Behavioral Sciences

Northwestern University

Acknowledgements

Hendricks Brown, Ph.D.   (Northwestern University)

Sanjay Mehrotra, Ph.D.   (Northwestern University)

Yun Li, Ph.D.   (CSR Audio Engineering Technology)

Cady Berkel, Ph.D.   (Arizona State University)

Willy Prado, Ph.D.   (University of Miami)

Hilda Pantin, Ph.D.   (University of Miami)

NORTHWESTERN
UNIVERSITY

REFERENCES

1. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas LA, et al. A computational future for preventing HIV in minority communities: How advanced technology can improve implementation of effective programs. *JAIDS Journal of Acquired Immune Deficiency Syndrome*. 2013.
2. Gallo CG, Pantin H, Villamar J, Prado G, Tapia M, Ogihara M, Cruden G, Brown CH. Blending Qualitative and Computational Linguistics Methods for Fidelity Assessment: Experience with the Familias Unidas Preventive Intervention. *Administration and Policy in Mental Health*. 2014.
3. Inoue, M., Ogihara, M., Hanada, R., & Furuyama, N. (2012). Gestural cue analysis in automated semantic miscommunication annotation. Multimedia tools and applications, 61(1), 7-20.
4. Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., & Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. Drug and Alcohol Dependence, 95, S5-S28.

NORTHWESTERN
UNIVERSITY

22