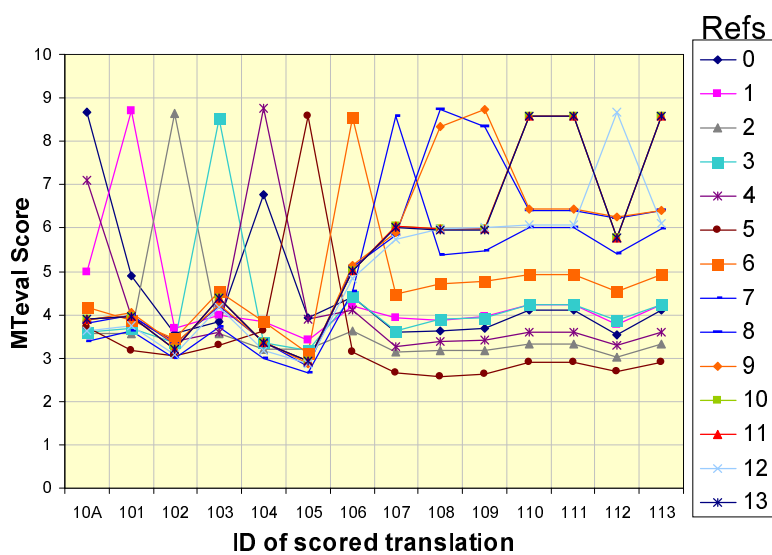


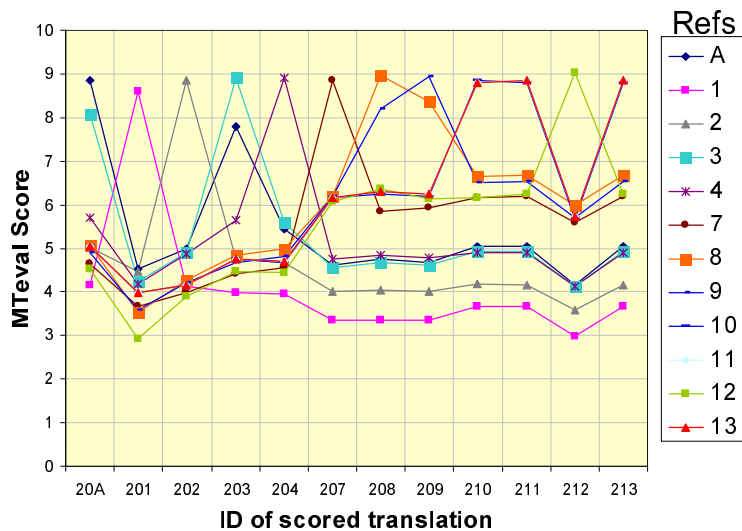
*Some Notes and Observations
on Scoring the Study Documents
using NIST's Mteval Scorer*

George Doddington, NIST

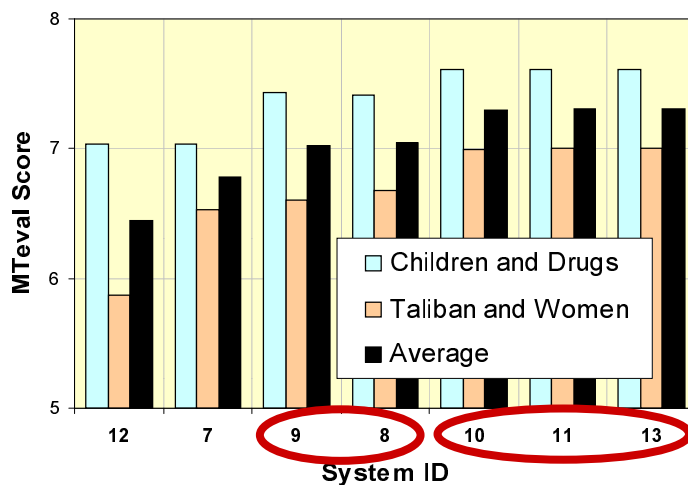
*Scoring "Children and Drugs"
using every translation as both a single ref and test*



Scoring "Taliban and Women" using every translation as both a single ref and test



Scores for Systems 7-13 Systems 1-6 Used for Reference



100 series : 10A as reference

• ID	SCORE	• ID	SCORE
• 20A	8.86	• 202	4.98
• 203	7.80	• 208	4.76
• 204	5.45	• 209	4.68
• 211	5.03	• 207	4.61
• 213	5.03	• 201	4.52
• 210	5.03	• 212	4.16

200 series : 20A as reference

• ID	SCORE	• ID	SCORE
• 10A	8.67	• 105	3.94
• 104	6.75	• 103	3.85
• 101	4.88	• 109	3.68
• 106	4.40	• 108	3.63
• 110	4.10	• 107	3.58
• 111	4.10	• 102	3.57
• 113	4.10	• 112	3.54

Notes

- Each translation was used in turn as reference
- The graphs lead to the following observations:
 - the following translations appear to be mere variants of each other: (20A and 203), (208 and 209) and (210, 211 and 213)
 - translations 207-213 all share a similarity that makes them categorically different from the other translations
 - similar conclusions hold for the 100 series (here, it is 10A and 104 that are variants of each other)