The following is a detailed description of the methods and results of the Rasch anslysis.

Methods

We examined data-model fit using information weighted (INFIT) and unweighted (OUTFIT) mean square values (MNSQ). INFIT and OUTFIT MNSQ are chi square statistics divided by their degrees of freedom and reported as ratios with an expected value of 1 and a range of 0 to infinity. These statistics provide an indication of the amount of useful information provided by an item. Values above 1 indicate more noise than expected by the model and values less than 1 indicate responses that are more deterministic than expected. Although there are no concrete rules about the acceptable thresholds for INFIT MNSQ and OUTFIT MNSQ, values between 0.5 and 1.5 are generally considered acceptable for use.[12,13] In addition, the Rasch reliability, an index of internal consistency similar to a Cronbach's alpha or KR-20, was determined.[14,15]

Results

One additional question from the KOOS Sport/Recreation subscale and four questions from the KOOS Quality of Life subscales were identified due to the low correlations between the KOOS,JR and these items. More than 67% of patients had preoperative responses of "moderate" or worse to all four questions from the Quality of Life subscale, thus all four items met the threshold for inclusion. While 71% of patients reported at least moderate difficulty when responding to the Sport/Recreation question regarding twisting and pivoting, consistent with the methods of Lyman et al.,[5] this question was not added to the $KOOS_{global}$ as the KOOS,JR already included a question related to pain during this specific task. Pre- and postoperative $KOOS_{global}$

scores were then calculated by combining the responses from the four Quality of Life questions with the seven KOOS,JR questions. Similar to the scoring of the KOOS,JR, the responses to the 11 $KOOS_{global}$ were summed to generate a raw score, and the raw score was then converted using a logit transformation (Table 1). The formula for converting raw to scaled scores was:

$$(-Logit + 6.1526) * 7.7701)$$

For the $KOOS_{global}$ score, the lowest possible raw score which would be indicative of no pain or dysfunction (0/44) was scaled to 100 and the highest possible raw score (44/44), which is indicative of extreme pain and dysfunction was scaled to 0.

Individual fit statistics for each item on the $KOOS_{global}$ are shown in Table 2. There were no items whose fit statistics exceeded previously established thresholds of .5 and 1.5. Overall, the data indicate sufficient fit to the model (Table 2). Person summary statistics for each instrument are shown in Table 3. Reliability, an index of internal consistency similar to a Cronbach's alpha or KR-20, was 0.89 for the $KOOS_{global}$.

Table 1. $KOOS_{global}$ raw scores to scaled scores conversion table.

Raw Score	Logit	SE	$KOOS_{global}$	Raw Score	Logit	SE	$KOOS_{global}$
0	-6.7173	1.885	100.000	23	0.184	0.3847	46.376
1	-5.3641	1.098	89.485	24	0.3328	0.3868	45.220
2	-4.4776	0.8251	82.597	25	0.4834	0.3895	44.049
3	-3.9035	0.7002	78.136	26	0.6363	0.3927	42.861
4	-3.469	0.6224	74.760	27	0.7919	0.3963	41.652
5	-3.1162	0.5683	72.019	28	0.9504	0.4002	40.421
6	-2.8161	0.5291	69.687	29	1.1123	0.4046	39.163
7	-2.5519	0.5001	67.634	30	1.2779	0.4093	37.876
8	-2.3132	0.478	65.779	31	1.4475	0.4145	36.558
9	-2.0934	0.4604	64.071	32	1.6217	0.4202	35.205
10	-1.8883	0.4458	62.478	33	1.8009	0.4268	33.812
11	-1.6952	0.4333	60.977	34	1.9863	0.4344	32.372
12	-1.5121	0.4225	59.555	35	2.1789	0.4436	30.875
13	-1.3377	0.4131	58.200	36	2.3806	0.4551	29.308
14	-1.1704	0.4051	56.900	37	2.5942	0.4699	27.648
15	-1.0091	0.3983	55.646	38	2.824	0.4898	25.863
16	-0.8527	0.3928	54.431	39	3.0769	0.5177	23.898
17	-0.7002	0.3886	53.246	40	3.3654	0.5591	21.656
18	-0.5504	0.3855	52.082	41	3.7129	0.6248	18.956
19	-0.4027	0.3835	50.935	42	4.1729	0.743	15.382
20	-0.256	0.3825	49.795	43	4.9159	1.0236	9.609
21	-0.1098	0.3824	48.659	44	6.1525	1.8382	0.000
22	0.0367	0.3832	47.520				

Table 2. Item Measures, Standard Errors, and Fit Statistics

	$KOOS_{global}$			
Item	**Measure**	**SE**	**INFIT MNSQ**	**OUTFIT MNSQ**
S6	.6902	.0266	1.08	1.18
P2	-.1286	.0226	1.19	1.15
P3	1.0442	.0268	1.02	1.04
P6	.6170	.0257	.97	.99
P9	1.5300	.0305	.93	.96
A3	1.5666	.0294	.87	.90
A5	.7641	.0257	.95	.98
Q1	-2.4345	.0220	.94	.93
Q2	-1.6873	.0208	1.29	1.38
Q3	-1.2673	.0211	1.06	1.01
Q4	-.6944	.0239	.65	.84

Measure is the mean item parameter calibration

S.E. is the mean standard error

INFIT MNSQ is an information weighted chi square statistic

OUTFIT MNSQ is an unweighted chi square statistic

Table 3. Person Summary Statistics (n=2020)

	Measure	S.E.	INFIT MNSQ	OUTFIT MNSQ	Rel.	Ext. (%)	
$KOOS_{global}$	-1.6853	.5059	1.00	1.01	.89	226	(5.6%)

Measure is the mean item parameter calibration

S.E. is the mean standard error

INFIT MNSQ is an information weighted chi square statistic

OUTFIT MNSQ is an unweighted chi square statistic

Rel. is Rasch reliability

Ext. is the numer and percentage of extreme scores

Three additional items were identified from the full version of the HOOS for potential inclusion in the $HOOS_{global}$. Less than 33% of patients had preoperative responses of "none" to question S1: "Do you feel grinding, hear clicking or any other type of noise from your hip?", question P1: "How often is your hip painful?", and question Q1: "How often are you aware of your hip?" S1 had an INFIT of 2.61 and OUTFIT of 2.91; an extreme violation of the established thresholds of .5 and 1.5. After removing the item, reliability increased from 0.90 to 0.92, which further suggests that the item was introducing only noise and not information. The raw

$HOOS_{global}$ score was then determined by summing the responses to the original six HOOS, JR items and questions P1 and Q1. The scaled $HOOS_{global}$ score was determined using the logit conversion table (Table 1). The formula for converting raw to $HOOS_{global}$ scaled scores was: (-Logit + 8.5004) * 5.8621. The $HOOS_{global}$ questionnaire and scoring instructions can be found in the Supplemental Files available on the journal's website.

The Individual fit statistics for each item on both the HOOS JR and $HOOS_{global}$ are shown in Table 2. There was one item (A5) whose OUTFIT exceed the previously established thresholds of .5 and 1.5; however, its INFIT did not, meaning that response patterns to this item are less predictable for people with calibrations further away from the item calibration. Overall, the data indicate sufficient fit to the model (Table 2). Person summary statistics for each instrument are shown in Table 3. Rasch reliability was 0.88 for the HOOS, JR and 0.92 for the $HOOS_{global}$.

Table 1. Raw score to scaled score conversion chart for the $HOOS_{global}$

Raw	Logit	SE	$HOOS_{global}$	Raw	Logit	SE	$HOOS_{global}$
0	-8.5583	1.9431	100.000	17	0.1401	0.5588	49.009
1	-7.0133	1.2371	90.943	18	0.4554	0.5644	47.161
2	-5.7641	1.0218	83.620	19	0.778	0.5719	45.269
3	-4.8853	0.8547	78.468	20	1.1104	0.5816	43.321
4	-4.2593	0.7358	74.799	21	1.4557	0.5942	41.297
5	-3.7727	0.6649	71.946	22	1.8183	0.6109	39.171
6	-3.3605	0.6225	69.530	23	2.2046	0.6333	36.907
7	-2.9904	0.5961	67.360	24	2.6238	0.6625	34.449
8	-2.6459	0.5789	65.341	25	3.0851	0.6961	31.745
9	-2.318	0.5672	63.419	26	3.5923	0.7271	28.772
10	-2.0011	0.5593	61.561	27	4.1387	0.7501	25.569
11	-1.6913	0.5542	59.745	28	4.7181	0.7737	22.172
12	-1.3859	0.5514	57.954	29	5.3466	0.8167	18.488
13	-1.0826	0.5503	56.177	30	6.0799	0.9067	14.189
14	-0.7797	0.5506	54.401	31	7.0879	1.1429	8.280
15	-0.4758	0.5521	52.619	32	8.5004	1.9031	0.000
16	-0.1697	0.5547	50.825				

Table 2. Item Measures, Standard Errors, and Fit Statistics

	HOOS, JR				$HOOS_{global}$			
Item	Measure	S.E.	INFIT MNSQ	OUTFIT MNSQ	Measure	SE	INFIT MNSQ	OUTFIT MNSQ
P5	-.3478	.0791	.92	.89	.7025	.0769	.93	.96
P10	-1.2067	.0775	.94	.93	-.0882	.0745	.86	.87
A3	.1986	.0975	.85	.90	1.2179	.0771	.90	1.21
A5	.1528	.0807	1.15	1.19	1.1689	.0782	1.21	1.76
A12	.5698	.0797	1.09	1.11	1.5856	.0774	1.12	1.30
A14	.6332	.0806	1.02	.97	1.6367	.0783	.98	.96
P1	---	---	---	---	-2.7101	.0723	.84	.90
Q1	---	---	---	---	-3.5134	.0717	1.00	1.12

Measure is the mean item parameter calibration

S.E. is the mean standard error

INFIT MNSQ is an information weighted chi square statistic

OUTFIT MNSQ is an unweighted chi square statistic

Table 3. Person Summary Statistics of the HOOS, JR and $HOOS_{global}$ (n=608)

Instrument	Measure	S.E.	INFIT MNSQ	OUTFIT MNSQ	Rel.	Ext. (%)
HOOS,JR	-2.1624	.7427	.99	1.00	.88	128 (21.5%)
$HOOS_{global}$	-1.7384	.6823	.97	1.09	.92	52 (8.9%)

Measure is the mean item parameter calibration

S.E. is the mean standard error

INFIT MNSQ is an information weighted chi square statistic

OUTFIT MNSQ is an unweighted chi square statistic

Rel. is Rasch reliability

Ext. is the numer and percentage of extreme scores